

# On the Origins and Control of Community Types in the Human Microbiome

Travis E. Gibson<sup>1,\*</sup>, Amir Bashan<sup>1,\*</sup>, Hong-Tai Cao<sup>2</sup>,  
Scott T. Weiss<sup>1</sup>, and Yang-Yu Liu<sup>1,3,†</sup>

<sup>1</sup> Channing Division of Network Medicine, Brigham and Women's Hospital,  
Harvard Medical School, Boston, Massachusetts 02115, USA

<sup>2</sup> Department of Electrical Engineering, University of Southern California,  
Los Angeles, California 90089, USA

<sup>3</sup> Center for Cancer Systems Biology, Dana-Farber Cancer Institute,  
Boston, Massachusetts 02115, USA

\* These authors contributed equally

† Corresponding Author: [yyl@channing.harvard.edu](mailto:yyl@channing.harvard.edu)

Microbiome-based stratification of healthy individuals into compositional categories, referred to as “community types”, holds promise for drastically improving personalized medicine. Despite this potential, the existence of community types and the degree of their distinctness have been highly debated. Here we adopted a dynamic systems approach and found that heterogeneity in the interspecific interactions or the presence of strongly interacting species is sufficient to explain community types, independent of the topology of the underlying ecological network. By controlling the presence or absence of these strongly interacting species we can steer the microbial ecosystem to any desired community type. This open-loop control strategy still holds even when the community types are not distinct but appear as dense regions within a continuous gradient. This finding can be used to develop viable therapeutic strategies for shifting the microbial composition to a healthy configuration.



# Contents

<b>1</b>	<b>Main Text</b>	<b>5</b>
1	Introduction . . . . .	5
2	Dynamic Model . . . . .	6
3	Metacommunity and Local Communities . . . . .	7
4	Origins of Community Types . . . . .	8
5	Discussion . . . . .	10
<b>2</b>	<b>Figures</b>	<b>13</b>
<b>3</b>	<b>Materials and Methods</b>	<b>19</b>
<b>4</b>	<b>Supplementary Figures</b>	<b>23</b>
<b>5</b>	<b>Supplementary Text</b>	<b>39</b>
1	Introduction . . . . .	39
2	Notation . . . . .	39
3	Random Variables and Random Matrices . . . . .	40
3.1	Primer on Random Variables . . . . .	40
3.2	Random Network Models . . . . .	42
3.3	Spectrum of Random Matrices . . . . .	43
4	Stability . . . . .	44
4.1	Primer on Stability . . . . .	44
4.2	Stability of Generalized Lotka Volterra Dynamics . . . . .	46
4.3	Stability in the presence of new species . . . . .	47
4.4	Robustness to disturbances. . . . .	49
4.5	Diagonal Stability of Random Matrices . . . . .	58
5	Clustering Analysis and Ordination Methods . . . . .	59
5.1	Distance and Metrics . . . . .	59
5.2	$k$ -Medoids . . . . .	59
5.3	Silhouette Value . . . . .	59
5.4	Variance Ratio Criterion and the Caliński-Harabasz Index . . . . .	60
5.5	Principle Coordinate Analysis . . . . .	60
5.6	Principle Component Analysis . . . . .	61
6	Modeling . . . . .	61
6.1	Universal Model . . . . .	61
6.2	Mutiple Set Model . . . . .	63
7	Simulation Results and Analysis . . . . .	63
7.1	Multiple Set Models . . . . .	63
7.2	Universal Model . . . . .	63
	<b>Bibliography</b>	<b>97</b>





# Chapter 1

## Main Text

### 1 Introduction

Rather than simple passengers in and on our bodies, commensal microorganisms have been shown to play key roles in our physiology and in the evolution of several chronic diseases [25, 81]. Many scientific advances have been made through the work of large-scale, consortium-driven metagenomic projects, such as the *Metagenomics of the Human Intestinal Tract* (MetaHIT) [71] and the *Human Microbiome Project* (HMP) [94, 95]. In particular, the HMP has analyzed the largest cohort and set of distinct, clinically relevant body habitats to date, in order to characterize the ecology of human-associated microbial communities [95]. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the translational applications of the human microbiome.

A recent study proposed that a healthy gut microbiome falls within one of three distinct community types, which the authors coined as “enterotypes” [3]. More specifically, the authors calculated the relative abundance profiles of microbiota at the genus level and then performed standard cluster analysis, finding three distinct clusters (enterotypes). Each enterotype is dominated by a particular genus (*Bacteroides*, *Prevotella*, or *Ruminococcus*) but not affected by gender, age, body mass index, or nationality of the host. These results suggest that enterotyping could be an efficient way to stratify healthy human individuals. The development of personalized microbiome-based therapies would then simplify to shifting an unhealthy microbiome to one of the distinct healthy configurations.

A meta-analysis, however, suggested that enterotypes, or in general community types, could be an artifact of the small sample size in [3] and what one should expect is a continuous gradient with dense regions rather than distinct clusters [52]. The level of discreteness or continuity of the community types remains unclear. Interestingly, samples in the dense regions of this gradient are either highly abundant or deficient in *Bacteroides* [52], indicating that community types could still emerge as the dense regions within a continuous gradient. Indeed, some recent work actually supports the notion of distinct community types [22, 41, 79, 102, 107].

We still lack consensus on the nature and origins of community types [4, 14, 32, 45, 51]. In principle the presence of community types can be explained by several different mechanisms. First, there may be true multi-stability, i.e. multiple stable states with all microbial species present in the same environment [57]. Those stable states are interior equilibrium points of the corresponding ecological dynamic system. Although this type of multi-stability has been well discussed in macro-ecological systems [17], its detection in host-associated microbial communities is rather difficult and has not been demonstrated experimentally, partially because any subtle differences in the environment can drive those microbial communities [32]. Second, there may be strong host heterogeneity, leading to host-specific microbial dynamics (parameterized by host-specific intra- and inter-species interactions). If those interactions, which serve as parameters of the host-associated microbial ecosystems, can be classified into distinct groups, then we can numerically demonstrate that distinct community types will naturally emerge (Supplementary Text Sec. 6.2 and 7.1). Yet, the presence of classifiable microbial dynamics has not been experimentally detected, presumably due to the lack of high-quality time-series data for a large number of

subjects. Moreover, the overwhelming success of *Fecal Microbiota Transplantation* (FMT) in treating *recurrent Clostridium Difficile Infection* (rCDI) actually implies that difference in conditions between individuals are unlikely the cause of community types [98, 105, 106].

Here we proposed a simple mechanism, without assuming multi-stability or host heterogeneity, to explain the origins of community types. In particular, using a dynamic systems approach, we studied compositional shift as a function of species collection and demonstrated that with heterogeneous interspecific interactions, a phenomenon often observed in macroecology [26, 27, 73], community types can naturally emerge. Interestingly, this result is independent of the topology of the underlying ecological network. To our knowledge, this is the first quantitative attempt to explore the analytical basis of community types. Furthermore, community types, even when they weakly exist, can be manipulated efficiently by controlling the *Strongly Interacting Species* (SISs) only.<sup>1</sup> This provides theoretical justification for translational applications of the human microbiome.

## 2 Dynamic Model

The human microbiome is a complex and dynamic ecosystem [35]. When modeling a dynamic system we should first decide how complex the model needs to be so as to capture the phenomenon of interest. A detailed model of the intestinal microbiome would include mechanistic interactions among cells, spatial structure of the human intestinal tract, as well as host-microbiome interactions [10, 68, 83, 97]. That level of detail however is not necessary for this study, because we are primarily interested in exploring the impact that any given species has on the abundance of other species. To achieve that, a population dynamics model such as the canonical *Generalized Lotka-Volterra* (GLV) model is sufficient [32, 76]. Indeed, GLV dynamics leveraging current metagenome data has already been used for predictive modeling of the intestinal microbiota [33, 62, 86].

Consider a collection of  $n$  species in a habitat with the population of species  $i$  at time  $t$  denoted as  $x_i(t)$ . The GLV model assumes that the species populations follow a set of ordinary differential equations

$$\dot{x}_i(t) = r_i x_i(t) + x_i(t) \sum_{j=1}^n a_{ij} x_j(t), \quad i = 1, \dots, n \quad (1)$$

where  $\dot{(\cdot)} = \frac{d}{dt}(\cdot)$ . Here  $r_i$  is the growth rate of species  $i$ ,  $a_{ij}$  (when  $i \neq j$ ) accounts for the impact that species  $j$  has on the population change of species  $i$ , and the terms  $a_{ii} x_i^2$  are adopted from Verhulst's logistic growth model [37]. By collecting the individual populations  $x_i(t)$  into a state vector  $x(t) = [x_1(t), \dots, x_n(t)]^T$ , Equation (1) can be represented in the compact form

$$\dot{x}(t) = \text{diag}(x(t)) (r + Ax(t)), \quad (2)$$

where  $r = [r_1, \dots, r_n]^T$  is a column vector of the growth rates,  $A = (a_{ij})$  is the interspecific interaction matrix, and  $\text{diag}$  generates a diagonal matrix from a vector. Hereafter we drop the explicit time dependence of  $x$ .

Next we discuss the notion of fixed point, or equivalently steady state, in the GLV dynamics. This notion is important in the context of the human microbiome, as the measurements taken of the relative abundance of intestinal microbiota in the aforementioned studies typically represent steady behavior [3, 95]. In other words, the intestinal microbiota is a relatively resilient ecosystem [61, 80], and until the next large perturbation (e.g. antibiotic administration or dramatic change in diet) is introduced, the system remains stable for months and possibly even years [13, 20, 31]. The fixed points of system (2) are those solutions  $x$  that satisfy  $\dot{x} = 0$ . The solution  $x = 0$  (i.e. all species have zero abundance) is a trivial steady state. The set of non-trivial steady states contains those solutions  $x^*$  such that  $r + Ax^* = 0$ . When the matrix  $A$  is invertible, it follows that the non-trivial steady state  $x^* = -A^{-1}r$  is unique [38].

Our study ultimately investigated the impact that different collections of microbial species have on their steady state abundances. In Box 1 we presented a detailed analysis

<sup>1</sup>In this paper we use the term species in the general context of ecology, i.e. a set of organisms adapted to a particular set of resources in the environment, rather than the lowest taxonomic rank. One could think of organizing microbes by genus or operational taxonomical units as well.

showing that if we introduce a new species into the ecosystem in (2), the shift of the steady state is proportional to the interaction strengths between the newly introduced species and the previously existing ones. Similarly, if two communities with the same dynamics differ by only one species, then it is the interaction strength of that species with regard to the rest of the community that dictates how far apart the steady states of the two communities will be. This analytical result indicates that heterogeneity of interspecific interactions could lead to the clustering of steady states, and hence the emergence of community types.

To systematically investigate how changes in species collection affect the steady state shift in the GLV dynamics, we assumed that two microbial species will interact in the same fashion regardless of the host. Otherwise, if the interactions are host specific and the dynamics are classifiable, we can show that distinct community types will emerge almost trivially (Supplementary Text Sec. 6.2 and 7.1).

### 3 Metacommunity and Local Communities

Consider a universal species pool, also referred to as a metacommunity [18], indexed by a set of integers  $\mathbf{S} = \{1, \dots, n\}$ , an  $n \times n$  matrix  $\mathbf{A}$  representing all possible pairwise interactions between species, and a vector  $\mathbf{r}$  of size  $n$  containing the growth rates for all the  $n$  species. The global parameters for the metacommunity are completely defined by the triple  $(\mathbf{S}, \mathbf{A}, \mathbf{r})$ . We consider  $q$  *Local Communities* (LCs), defined by sets  $S^{[\nu]}$  that are subsets of  $\mathbf{S}$ , denoting the species present in  $\text{LC}_\nu$  with  $\nu = 1, \dots, q$ . This modeling procedure is inspired by the fact that alternative community assembly scenarios could give rise to the compositional variations observed in the human microbiome [18]. These LCs represent microbial communities in the same body site across different subjects. For simplicity, we assume that each LC contains only  $p$  species ( $p \leq n$ ), randomly selected from the metacommunity.

The GLV dynamics for each LC is given by

$$\text{LC}_\nu : \quad \dot{x}^{[\nu]}(t) = \text{diag}\left(x^{[\nu]}(t)\right) \left(r^{[\nu]} + A^{[\nu]}x^{[\nu]}(t)\right), \quad (3)$$

where the LC specific interaction matrix and growth vector are defined as  $A^{[\nu]} = \mathbf{A}_{S^{[\nu]}, S^{[\nu]}}$  and  $r^{[\nu]} = \mathbf{r}_{S^{[\nu]}}$ , respectively. That is,  $A^{[\nu]}$  is obtained from  $\mathbf{A}$  by only taking the rows and columns of  $\mathbf{A}$  that are contained in the set  $S^{[\nu]}$ . A similar procedure is performed in order to obtain  $r^{[\nu]}$ . Finally for each  $x^{[\nu]}$  there is a corresponding  $\mathbf{x}^{[\nu]} \in \mathbb{R}^n$  that has the abundances for species  $S^{[\nu]}$  of  $\text{LC}_\nu$  in the context of the metacommunity species pool  $\mathbf{S}$ .

To reveal the origins of community types in the human microbiome, we decomposed the universal interaction matrix as

$$\mathbf{A} = \mathbf{N}\mathbf{H} \circ \mathbf{G}s, \quad (4)$$

which contains four components. (i)  $\mathbf{N} \in \mathbb{R}^{n \times n}$  is the nominal interspecific interaction matrix where each element is sampled from a normal distribution with mean 0 and variance  $\sigma^2$ , i.e.  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, \sigma^2)$ . (ii)  $\mathbf{H} \in \mathbb{R}^{n \times n}$  is a diagonal matrix that captures the overall interaction strength heterogeneity of different species. When studying the impact of interaction strength heterogeneity the diagonal elements of  $\mathbf{H}$  will be drawn from a power-law distribution with exponent  $-\alpha$ , i.e.  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$ , which are subsequently normalized so that the mean of the diagonal elements is equal to 1. This is to ensure that the average interaction strength is bounded. For studies that do not involve interaction strength heterogeneity  $\mathbf{H}$  is simply the identity matrix. (iii)  $\mathbf{G} \in \mathbb{R}^{n \times n}$  is the adjacency matrix of the underlying ecological network:  $[\mathbf{G}]_{ij} = 1$  if species  $i$  is affected by the presence of species  $j$  and 0 otherwise. For details on the construction of  $\mathbf{G}$  for different network topologies see Supplementary Text Sec. 3.2.2. Note that the *Hadamard product* ( $\circ$ ) between  $\mathbf{H}$  and  $\mathbf{G}$  represents element-wise matrix multiplication. (iv) The last component  $s$  is simply a scaling factor between 0 and 1. Finally, we set  $[\mathbf{A}]_{ii} = -1$ . The presence of the scaling factor  $s$  and setting the diagonal elements of  $\mathbf{A}$  to  $-1$  are to ensure an asymptotic stability condition for the GLV dynamics (Supplementary Text Sec. 4.2, 4.3.3, and 4.5). The elements in the global growth rate vector  $\mathbf{r}$  are taken from the uniform distribution,  $[\mathbf{r}]_i \sim \mathcal{U}(0, 1)$ . Details concerning the distribution  $\mathcal{N}$ ,  $\mathcal{P}$  and  $\mathcal{U}$  can be found in Supplementary Text Sec. 3.1.1.

## 4 Origins of Community Types

We first studied the role of interspecific interaction strength heterogeneity on the emergence of community types. In order to achieve this, we chose the complete graph topology, i.e. each species interacts with all other species. This eliminates any *structural* heterogeneity. The nominal interaction strengths were taken from a normal distribution  $\mathcal{N}(0, 1)$ , the scaling component was set to  $s = 0.7$ , and the interaction strength heterogeneity was varied from low heterogeneity ( $\alpha = 7$ ) to a high level of heterogeneity ( $\alpha = 1.01$ ). Figure 1 displays the distributions of the diagonal elements of the interaction heterogeneity matrix  $\mathbf{H}$  at various heterogeneity levels. For each level of heterogeneity we constructed 500 LCs, each with 80 species randomly drawn from a metacommunity of 100 species. Figure 1b illustrates the global interaction matrix  $\mathbf{A}$  as a weighted network. With low heterogeneity all the link weights are of the same order of magnitude. As the heterogeneity increases fewer nodes contain highly weighted links, until there is only one node with highly weighted links when  $\alpha = 1.01$ . These nodes with highly weighted links correspond to SISs.

Figure 1c presents the results of *Principle Coordinates Analysis* (PCoA) of the steady states associated with the 500 different LCs as a function of  $\alpha$ . For low interaction heterogeneity ( $\alpha = 7$ ) the classical clustering measure, Silhouette Index, is less than 0.1, suggesting a lack of clustering in the data. As the heterogeneity increases the steady states can be seen to separate in the first two principle coordinate axes. At one point ( $\alpha = 2.0$ ) three clusters is the optimal number of clusters. Then as  $\alpha$  continues to decrease the optimal number of clusters becomes two. The fact that there are three clusters when  $\alpha = 2.0$  is not special, as a different number of optimal clusters can be observed with different model parameters or different clustering measures (see Supplementary Text Sec. 7.2) [52]. While the precise number of clusters is not important here, what is important is the fact that the degree of interaction strength heterogeneity controls the degree to which the clusters appear to be distinct. For low levels of interaction strength heterogeneity the clusters appear to be more like dense regions within a continuous gradient. As the heterogeneity increases, the clusters become more distinct. Indeed, having two clusters for  $\alpha = 1.01$  is to be expected, because one of the clusters is associated with all the LCs that contain the single SIS, and the other LCs that do not contain the single SIS constitute the other cluster.

The overall trend observed in Figure 1c is unaffected if the complete graph is replaced by an *Erdős-Rényi* (ER) random graph, or if the total number of LCs is increased (Supplementary Figures S1 and S2). The result is also generally unaffected by the specifics of the nominal distribution (Supplementary Text Sec. 7.2.1), the mean degree of the ER graph (Supplementary Text Sec. 7.2.2), or the number of species in the LCs (Supplementary Text Sec. 7.2.3). Of course, each LC can be invaded by other species that are currently absent. If this migration occurs relatively fast, then all LCs will converge to roughly the same species collection and the clustering will disappear. Hence in our modeling approach we have to assume that the migration occurs at a relatively slow time scale, and the time interval between species invasions is too long to disrupt the clustering. We also note that if heterogeneous interactions are placed at random in the network the clustering of steady states does not arise (Supplementary Figure S3). Our results are also robust (in the control theoretical sense) to stochasticity and the migration of existing species [34]. Robustness to migration is illustrated in Supplementary Figures S4 and S5, and robustness to stochastic disturbances is illustrated in Supplementary Figures S6-S8 (see Supplementary Text Sec. 4.4 for analytical robustness results).

We can explain the above results as follows: for low interaction strength heterogeneity all of the matrices  $A^{[v]}$  are very similar. In other words, despite containing different sets of species, all the LCs have very similar dynamics. Thus, clustering of steady states is not to be expected. As the heterogeneity of interaction strength increases, however, some of the LCs will have species that are associated with the highly weighted columns in  $\mathbf{A}$ , i.e. the SISs. Figure 2 presents a detailed analysis of the most abundant (dominating) species in each of the three clusters (community types) in Figure 1c for  $\alpha = 2$  and  $\alpha = 1.6$ , along with the abundances of the SISs within each cluster. It is clear that for different clusters their dominating species are different, consistent with the empirical finding that each enterotype is dominated by a different genus [3]. The SISs that are present in each cluster also vary. For instance with  $\alpha = 1.6$  all LCs in the blue cluster contain SISs number 23 and 81, and

none have species 60 or 51. For the orange cluster it is the opposite scenario. All of the LCs in the orange cluster contain SISs 60 and 51, and do not contain species 23 or 81. Most of the LCs in the yellow cluster contain SISs 23 and 51. Hence, each community type is well characterized by a unique combination of SISs. Note that none of the SISs are dominating species. These findings, along with the analysis in Box 1, suggest that heterogeneity in interaction strengths or the presence of SISs leads to the clustering of steady states, i.e. the emergence of community types.

We then studied the impact of structural heterogeneity on community types. Four different scenarios are illustrated in Figure 3: (a) a complete graph topology as in Figure 1; (b) an ER random graph as in Supplementary Figure 1; (c) a power-law out-degree network; (d) a power-law out-degree network with *no* interaction strength heterogeneity. Figures 3a, 3b and 3c support the main result shown in Figure 1, i.e. increasing interaction strength heterogeneity leads to the emergence of distinct community types. Figure 3d displays rather unexpected results as it suggests that structural heterogeneity alone does not lead to distinct community types. It is only with the inclusion of interaction strength heterogeneity that structurally heterogeneous microbial ecosystems can display strong clustering in their steady states as shown in Figure 3c. This result is rather surprising, because structural heterogeneity is observed in many real-world complex networks [1, 6, 58] and has been shown to affect many dynamical processes over complex networks [60, 72, 75].

Note that in the preparation of Figure 3 the steady state abundances were normalized to get relative abundances of the species and the Jensen-Shannon distance metric was used for clustering analysis [39]. The trends discussed above also hold when, instead of the Silhouette Index, the Variance Ratio Criterion is used as the clustering measure, or the Euclidean distance is used for clustering, or when absolute abundances are analyzed along with the Euclidean distance being used (Supplementary Figures S9, S10, and S11). Supplementary Figure S11 correlates to the analytical results in Box 1, where absolute abundances and the Euclidean distance are implicitly used.

## Control of Community Types

With the knowledge that each community type can be associated with a specific collection of SISs, we tested the hypothesis that a local community could be steered to a desired community type by controlling the combination of SISs only. Our results for three different scenarios are shown in Figure 4a for  $\alpha = 1.6$ . The local community that was controlled in each scenario is shown in magenta and is denoted  $LC^*$ , which initially belongs to the blue cluster. For Scenario 1,  $LC^*$  had the SISs 23 and 81 removed, with species 60 and 51 simultaneously introduced with random initial abundances drawn from  $\mathcal{U}(0, 1)$ . Recall that species 60 and 51 are the SISs present in the orange cluster. This swap of SISs shifts  $LC^*$  to a slightly different state (green dot) within the blue cluster. The GLV dynamics were then simulated and the trajectory goes from the blue cluster to the orange cluster. This result was independent of the initial condition of species 60 and 51 (Fig. 4b). This open-loop control of the community type by manipulating a set of SISs also works at lower levels of heterogeneity (Fig. 4c and 4d). Here we use the term open-loop to contrast closed-loop control where inputs are designed with feedback so as to continuously correct the system of interest. These findings imply that the SISs, despite their low abundances, can be used to effectively control a microbial community to a desired community type.

In Scenario 2 we tested if the same result could be obtained by removing the six most abundant species from  $LC^*$  and introducing the six most abundant species from the orange cluster at exactly the same abundance level as an arbitrary local community in the orange cluster. The state after this dominating species swap (red dot) starts close to the orange cluster, because the six most abundant species from a local community in that cluster were copied. The trajectory does not ultimately converge near the orange cluster, but goes toward the blue cluster instead. The trajectory, however, does not ultimately converge in the blue cluster because it does not contain any of the most abundant species present in the blue cluster.

In scenario 3 we explored how the open-loop control methodology just presented could also be used to conceptually justify the success of FMT in treating patients with rCDI [98, 105, 106]. This scenario begins by removing 20 species from  $LC^*$  (the top two SISs and

18 of the most abundant spaces) so as to emulate the effect of broad-spectrum antibiotics, resulting in an altered community (blue dot). Then the GLV dynamics were simulated and the local community converged to a new steady state (black dot), representing the CDI state. To emulate an oral capsule FMT 1% of the species abundances from an arbitrary LC in the orange cluster, i.e. the donor, was added to the CDI state, resulting in a slightly altered community (gray dot). The GLV dynamics were then simulated until the final steady state was reached (white dot). As expected the post-FMT steady state is in the orange cluster, the same cluster that is associated with the donor's LC. Note that if during the FMT the SISs in the donor's LC were not transplanted then the patient's post-FMT steady state does not converge in the orange cluster (Supplementary Figure S12).

The above results indicate that the presence of SISs simplifies the open-loop control design. However, the existence of community types is not a prerequisite for deploying this control methodology. The possibility for open-loop control of the human microbiome will likely be body site specific. Our work focused on the gut specifically because of the fact that this microbial community is very likely dominated by microbe-microbe and/or host-microbe interactions, rather than external disturbances. It is yet to be determined what factors drive the dynamics in other body sites.

## 5 Discussion

In this work we studied compositional shift as a function of species collection using a dynamic systems approach, aiming to offer a possible mechanism for the origins of community types. We found that the presence of interaction strength heterogeneity or SISs is sufficient to explain the emergence of community types in the human microbiome, independent of the topology of the underlying ecological network. The presence of heterogeneity in the interspecific interaction strengths in natural communities has been well studied in macroecology [26, 27, 64, 73]. Extensive studies are still required to explore this interesting direction in the human microbiome. While preliminary analysis is promising, all existing temporal metagenomic datasets are simply not sufficiently rich to infer the interspecific interaction strengths among all of the microbes present in and on our bodies [32] even at the genus level, let alone the species level. Recent studies have tried to overcome this issue by only investigating the interactions between the most abundant species [33]. Our results, however, suggest that SISs need not be the most abundant ones and can still play an important role in shaping the steady states of microbial ecosystems. Ignoring the lack of sufficient richness, system identification analysis with regularization and cross-validation [11, 86] of the largest temporal metagenomic dataset to date [13] does not disprove the existence of SISs. To the contrary, it supports this assertion (see Supplementary Figure S13). Permutation of the time series however also results in the identification of interaction strength heterogeneity (see Supplementary Figures S14 and S15). Hence, the presence of SISs needs to be systematically studied with novel system identification methods and perhaps further validated with co-culture experiments [32]. For example, we could first use metabolic network models to predict levels of competition and complementarity among species [56], which could then be used as prior information to further improve system identification [2].

Note that our notion of SIS is fundamentally different from that of keystone species, which are typically understood as species that have a disproportionate deleterious effect (relative to its abundance) on the community upon their removal [74]. One can apply a brute-force leave-one-out strategy to evaluate the “degree of keystone-ness” of any species in a given community network [7]. Even without any interaction strength heterogeneity, a given community may still have a few keystone species. The SISs defined in this work are those species that have very strong impacts (either positive or negative) on the species that they directly interact with. The presence of SISs requires the presence of interaction strength heterogeneity. We emphasize that an SIS is not necessarily a keystone species. In fact, without any special structure embedded in the interaction matrix (and hence the ecological network), there is no reason why the removal of any SIS would cause mass extinction. It does have a profound impact on the steady-state shift, which is exactly what we expected from our analytical results presented in Box 1.

Our findings also have important implications as we move forward with developing microbiome-based therapies, whether it be through drastic diet changes, FMT, drugs, or



even engineered microbes [53, 55, 65, 85, 89, 103, 104]. Indeed, our results suggest that a few strongly interacting microbes can determine the steady state landscape of the whole microbial community. Therefore, it may be possible to control the microbiome efficiently by controlling the collection of SISs present in a patient’s gut. Finer control may be possible through the engineering of microbes. This will involve a detailed mechanistic understanding of the metabolic pathways associated with the microbes of interest. As discussed in Box 1, given a new steady state of interest, the parameters  $b, c, d, s$  could be chosen such that the new steady state is feasible and stable (Supplementary Text Sec. 4.3.1). Then, with the knowledge of the appropriate parameters  $b, c, d, s$  it would be possible to introduce a known microbe with those characteristics or engineer one to have the desired properties. We emphasize that the stability and control of the microbial ecosystem must be studied at the macroscopic scale using a systems and control theoretic approach. This is similar to what is carried out in aerospace applications. The design of wings and control surfaces for an aircraft incorporate sophisticated fluid dynamic models. The control algorithms for planes however are often derived from simple linearized reduced order dynamic models where linear control techniques can be easily deployed [87]. Taken together, our results indicate that the origins and control of community types in the human microbiome can be explored analytically if we combine the tools of dynamic systems and control theory, opening new avenues to translational applications of the human microbiome.

## Acknowledgements

We thank George Weinstock, Curtis Huttenhower, Rob Knight, Domitilla Del Vecchio, and Doug Lauffenburger for helpful discussions. Special thanks to Aimee Milliken for a careful reading of the text. This work was partially supported by the John Templeton Foundation (award number 51977) and National Institutes of Health (R01 HL091528).

## Contributions

Y-YL conceived the project. All authors performed research. AB performed initial numerical calculations. TEG performed systematic analytical and numerical calculations. HTC and TEG performed the system identification. TEG and Y-YL wrote the manuscript. AB and STW edited the manuscript.





## Chapter 2

## Figures

### Box 1. Steady State Shift in the Generalized Lotka-Volterra Model

*Impact of one new species:* Consider the addition of one new species to the system described by (2), so that now we have  $m = n + 1$  species, with the abundance of the  $i$ -th species denoted as  $z_i(t)$ ,  $i = 1, \dots, m$ . The new  $m$ -dimensional state vector  $z(t) = [z_1(t), z_2(t), \dots, z_m(t)]^T$  evolves as

$$\dot{z}(t) = \text{diag}(z(t))(g + Fz(t)), \quad (\text{B1})$$

where

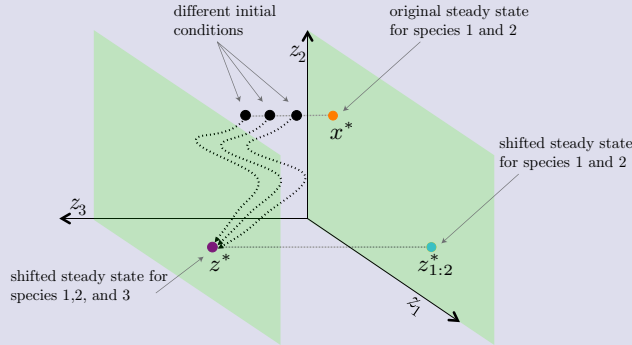
$$g = \begin{bmatrix} r \\ s \end{bmatrix}, \quad \text{and} \quad F = \begin{bmatrix} A & b \\ c^T & d \end{bmatrix}.$$

Note here that  $A$  and  $r$  are as defined in (2) and we have only introduced four new elements  $s, b, c, d$ . The scalar element  $s$  represents the growth rate of the additional species  $z_m$ . The  $n$ -dimensional vector  $b$  represents the impact that species  $z_m$  has on the first  $n$  species  $z_{1:n}$ , which also correspond to the  $n$  species in the original state vector  $x$ . The scalar element  $d$  is the Verhulst term for the new species  $z_m$ . Finally, the  $n$ -dimensional vector  $c$  represents the effect that  $z_{1:n}$  have on the dynamics of the  $m$ -th state  $z_m$ .

Given the dynamics in (2) and (B1), with the assumption that the matrix  $A$  is full rank, we can show that the difference between the original steady state  $x^*$  and the shifted steady state of the same  $n$  species, denoted as  $z_{1:n}^*$ , satisfies the following equality

$$z_{1:n}^* - x^* = -A^{-1}bz_m^*, \quad (\text{B2})$$

where  $z_m^*$  is the steady state value of the newly added species. Given that  $A$  is the same in (2) and (B1), the shift in the steady state of the original  $n$  species is bilinear in terms of the vector  $b$  and the steady state value of the newly added  $m$ -th species  $z_m^*$ .

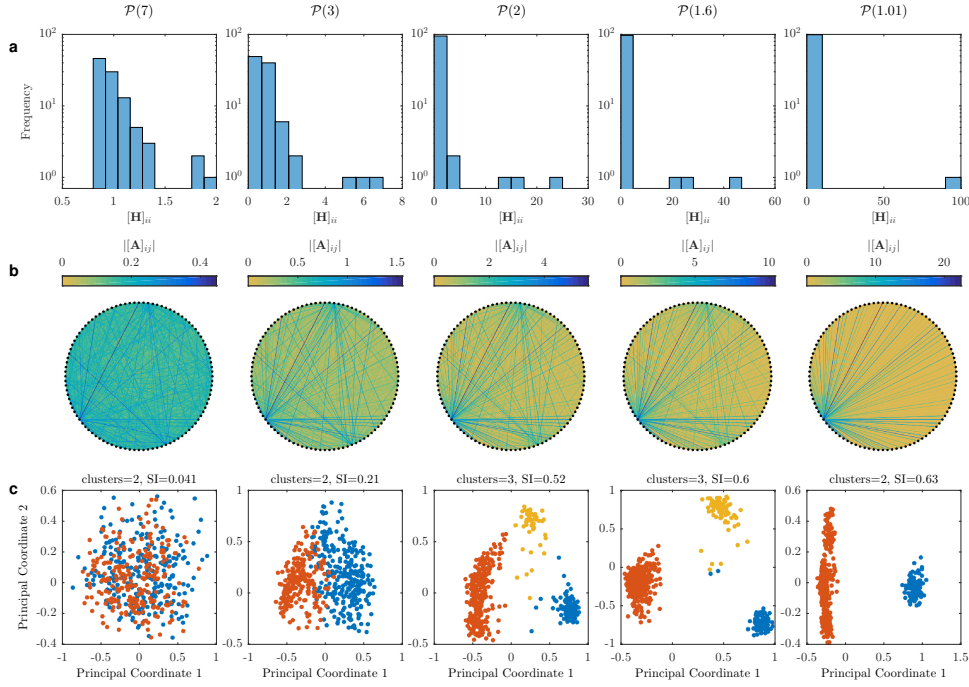


*Open-loop Control of Steady State:* If there exists a diagonal matrix  $P > 0$  such that  $A^T P + PA < 0$  then for any  $z^*$  there exists  $b, c, d, s$  such that the original steady state  $x^*$  of (2) can be steered to  $z^*$  and furthermore this  $z^*$  can be made to be uniformly asymptotically stable (Supplementary Text Sec. 4.3.1, Theorem 8).

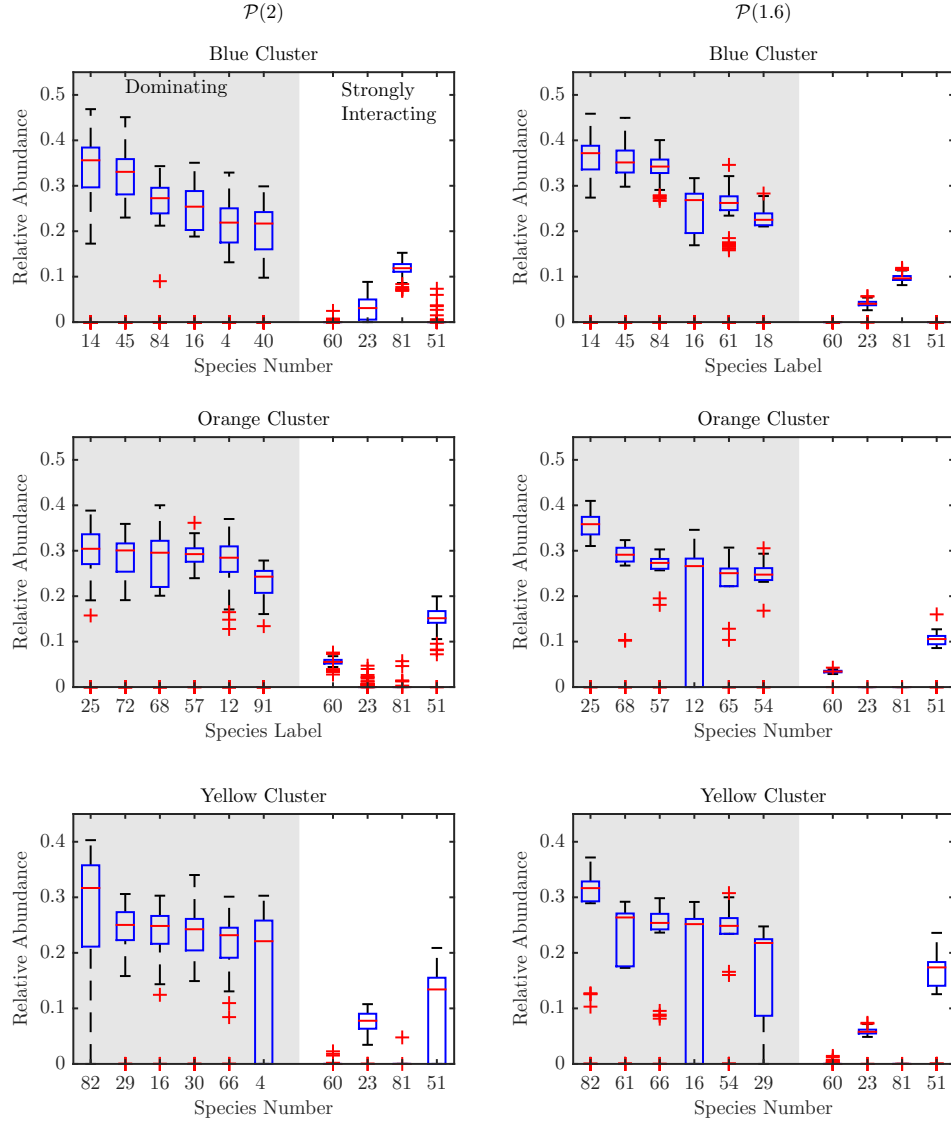
*Impact of multiple non-common species:* Consider two systems with different collections of species. Let  $\bar{z}^*$  be the steady state of system 1 and  $\hat{z}^*$  be the steady state of system 2. Assuming that the systems share  $n$  species, then we can just apply the results from (B2) recursively to calculate the difference between the abundances of the common species as

$$\bar{z}_{1:n}^* - \hat{z}_{1:n}^* = A^{-1} \left( \sum_{i \in \bar{M}} \bar{b}_i z_i^* - \sum_{i \in \hat{M}} \hat{b}_i \hat{z}_i^* \right)$$

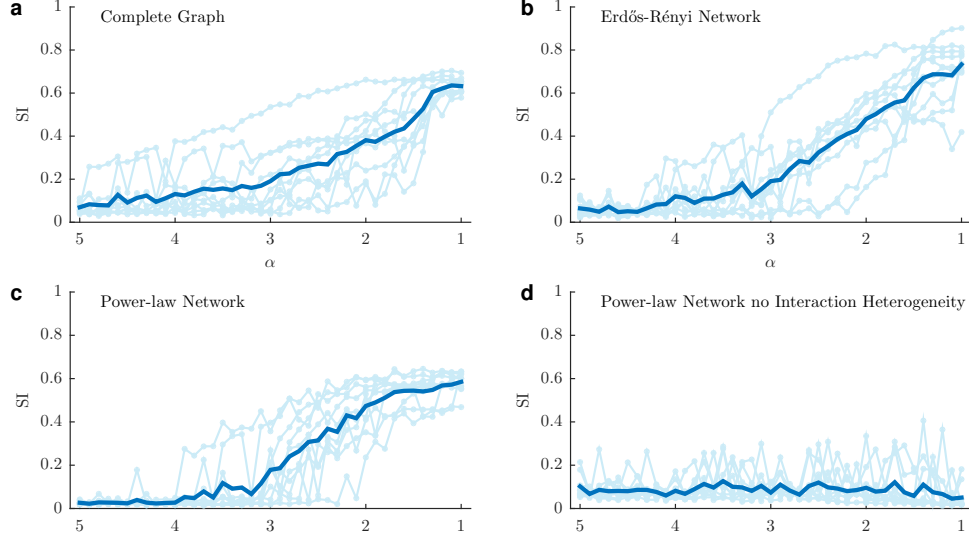
where  $\bar{M}$  and  $\hat{M}$  are the indices of the non-common species of system 1 and system 2 respectively.



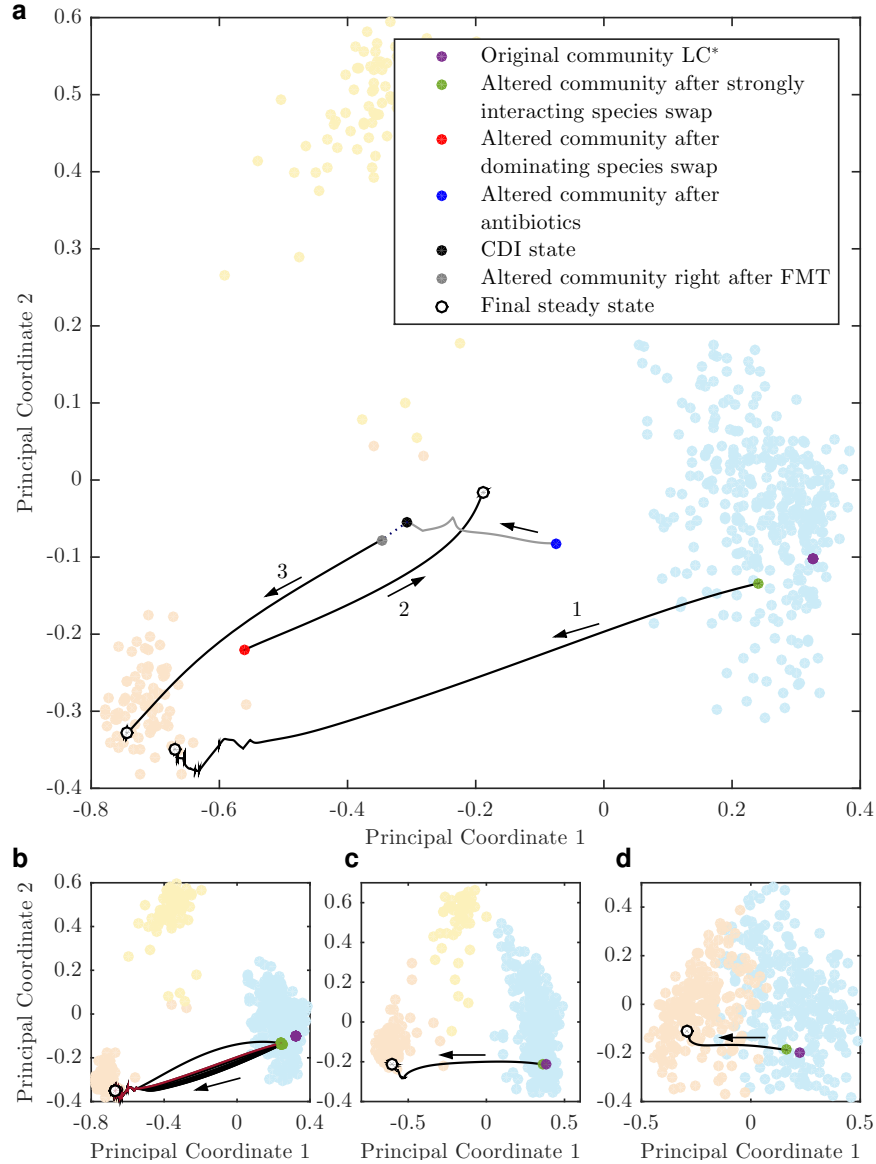
**Figure 1:** *Impact of interaction strength heterogeneity on the distinctness of community types.* A total of  $q = 500$  local communities, each with  $p = 80$  species randomly drawn from a universal pool of  $n = 100$  species. The nominal components were drawn from  $\mathcal{N}(0, 1)$ , the interaction heterogeneity matrix elements were taken from  $\mathcal{P}(\alpha)$  and  $\alpha$  is varied with the set of values  $\{7, 3, 2, 1.6, 1.01\}$  for each column in the figure. The topology component  $\mathbf{G}$  has all elements equal to 1, giving a complete graph. The scaling factor was set at  $s = 0.07$ . (a) Histogram of the diagonal elements of the heterogeneity matrix  $\mathbf{H}$ . (b) Visualization of the universal interaction matrix  $\mathbf{A}$  as a weighted adjacency matrix of a digraph. (c) Principle coordinate analyses of the normalized steady state for each local community using the Jensen-Shannon distance. The Silhouette Index and optimal number of clusters are denoted. Further details can be found in the Methods Section.



**Figure 2:** Comparison of dominating species to SISs in different community types (clusters). The relative abundances of the six most abundant species from each of the three clusters in Figure 1c for  $\alpha = 2$  and  $\alpha = 1.6$  are compared to that of the four species with the largest interaction strengths (60, 23, 81, and 51).



**Figure 3:** *Impact of network structure on the distinctness of community types.* For each type of network structure 10 different Universal Triples  $(\mathbf{S}, \mathbf{A}, \mathbf{r})$  with  $n = 100$  species and  $q = 500$  local communities of size  $p = 80$  were generated with results shown in the lighter color and averaged results shown in bold. (a) *Complete graph.* Same study as in Figure 2 with  $\alpha \in [5, 1]$ . (b) *Erdős-Rényi network (digraph)*  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$ ,  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$  where  $\alpha \in (1, 5]$ , Probability  $[\mathbf{G}]_{ij} = 1$  is 0.1, i.e. a mean in(out)-degree of 10, and scaling factor  $s = 1/\sqrt{10}$ . (c) *Power-law out-degree network*  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$ ,  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$ ,  $\mathbf{G}$  is the adjacency matrix for a digraph with out-degree having a power-law distribution  $\mathcal{P}(\alpha)$ . The high-degree nodes have the largest interaction scaling. (d) *Power-law out-degree network, no interactions strength heterogeneity*  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$ ,  $\mathbf{H}$  is the identity matrix,  $\mathbf{G}$  is the adjacency matrix for a digraph with out-degree having a distribution  $\mathcal{P}(\alpha)$ . Further details can be found in the Methods Section.



**Figure 4:** Open-loop control of the human microbiome. (a) Background of clustering analysis for  $\alpha = 1.6$  from Figure 1, but with Euclidean distance used so that a projection matrix could be found to show the trajectories in the 2D principle coordinate plane (Supplementary Text Sec. 5.6). We aim to steer a local community (denoted as LC\*, shown in magenta) in the blue cluster to the orange cluster. Three different scenarios are presented, per the three numbers above the arrows. Scenario 1: SISs swap. The SISs (23 and 81) of LC\* were replaced by the SISs present in the orange cluster (60 and 51). The initial abundances of species 60 and 51 were drawn from  $\mathcal{U}(0, 1)$ , resulting in the green dot, and the GLV dynamics were simulated. Scenario 2: dominating species swap. The six most abundant species in LC\* were removed and replaced by the six most abundant species from a local community in the orange cluster, with the initial condition after the switch of species shown as the red dot, and the dynamics were simulated until steady state was reached. Scenario 3: Fecal Microbiota Transplantation (FMT). The two SISs and 18 of the most abundant species (for a total of 20) were removed from LC\* with the initial condition shown in blue (post-antibiotic state). Then the GLV dynamics were simulated (gray line) and the system converged to the black dot (CDI state). Then 1% of the steady abundances from an arbitrary LC in the orange cluster were added to the CDI state (gray dot, emulating oral capsule FMT) and the dynamics were then simulated until steady state was reached. (b) The SISs swap process was repeated ten times, each time the initial abundances of species 60 and 51 were randomly drawn from  $\mathcal{U}(0, 1)$ . Nine of the simulations are shown in black and the simulation that pertains to Figure 4a is shown in maroon. (c) The same analysis as for Figure 4a, in terms of SISs swap, but for  $\alpha = 2$ . (d) The same analysis as for Figure 4a, in terms of SISs swap, but for  $\alpha = 3$ .



## Chapter 3

# Materials and Methods

The methods section begins with a toy example to illustrate the construction of the universal interaction matrix  $\mathbf{A} = \mathbf{N}\mathbf{H} \circ \mathbf{G}$  in (4), where

$$\begin{aligned}
 \text{steps: } (i) \quad & \mathbf{N} = \begin{bmatrix} 0 & 0.2 & 0.4 & -0.1 \\ 0.7 & 0 & 0.3 & 0.4 \\ -0.1 & 0.7 & 0 & 0.1 \\ -0.3 & -0.2 & 0.4 & 0 \end{bmatrix} \\
 (ii) \quad & \mathbf{H} = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0.4 \end{bmatrix} \\
 (iii) \quad & \mathbf{G} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\
 (iv) \quad & s=1 \\
 (v) \quad & [\mathbf{A}]_{ii} = -1 \\
 \text{final result:} \quad & \mathbf{A} = \begin{bmatrix} -1 & 0.04 & 0.08 & -0.04 \\ 7 & -1 & 0.06 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 0 & 0.08 & -1 \end{bmatrix}
 \end{aligned}$$

Given that  $\mathbf{H}$  is diagonal, it scales the columns of  $\mathbf{N}$ . If one thinks of  $\mathbf{A}$  as the adjacency matrix of a digraph, then  $\mathbf{H}$  scales all of the edges leaving a node. Thus one can consider  $\mathbf{H}$  as controlling the interaction strength heterogeneity of  $\mathbf{A}$ . Given the Hadamard product between  $\mathbf{H}$  and  $\mathbf{G}$ , the off-diagonal elements of  $\mathbf{G}$  that are zero will result in the corresponding off-diagonal elements of  $\mathbf{A}$  being zero as well.

In the first study (Figure 1), to explore the impact of interaction heterogeneity on steady state shift, we varied the exponent  $-\alpha$  of the power-law distribution of  $[\mathbf{H}]_{ii}$  to generate five different universal interaction matrices  $\mathbf{A}$  of dimension  $100 \times 100$ . For each universal interaction matrix  $\mathbf{A}$ , the nominal component  $\mathbf{N}$  consists of independent and identically distributed elements sampled from a normal distribution  $\mathcal{N}(0, 1)$ . The topology for this study was a complete graph and thus all the elements in  $\mathbf{G}$  are equal to 1. The heterogeneity element  $\mathbf{H}$  is constructed in two steps. First, five different vectors  $\bar{h}(\alpha) \in \mathbb{R}^{100}$  are constructed where each element is sampled from a power-law distribution  $\mathcal{P}(\alpha)$  for  $\alpha \in \{7, 3, 1.6, 1.2, 1.01\}$ . Then, each of the  $\bar{h}(\alpha)$  is normalized to have a mean of 1,  $h = \bar{h}/\text{mean}(\bar{h})$ . Finally the heterogeneity matrix is defined as  $\mathbf{H} = \text{diag}(h)$ . For this study  $s = 0.07$ , ensuring uniform asymptotic stability for the case of low heterogeneity (see Supplementary Text Theorem 17). The final step in the construction of  $\mathbf{A}$  is to set the diagonal elements to  $-1$ .

For each  $\alpha$  the following simulation steps were taken. There are a total of 100 species,  $\mathbf{S} = \{1, 2, \dots, 100\}$ , in the metacommunity, and each of the 500 local communities contains 80 species, randomly chosen from  $\mathbf{S}$ . The MATLAB command used to perform this step is `randperm`. The initial condition for each of the 500 local communities,  $x^{[\nu]}(0)$ ,

were sampled from  $\mathcal{U}(0, 1)$ . The dynamics were then simulated for 10 seconds using the MATLAB command `ode45`. If any of the 500 simulations crashed due to instability or if the norm of the terminal discrete time derivative was greater than 0.01 then that local community was excluded from the rest of the study. Those simulations that finished without crashing and with small terminal discrete time derivative were deemed steady. Less than 1% of simulations were deemed unstable in the preparation of Figure 1.

The networks presented in the second row of Figure 2 were constructed by considering  $\mathbf{A}$  as the weighted adjacency matrix of the network. Note that arrows showing directionality and self loops were suppressed. The links were color coded in proportion to the absolute value of the entries in  $\mathbf{A}$ .

For the last row of Figure 1 a clustering analysis was performed. For each  $\alpha$  the steady state abundances of the 500 local communities were normalized so that we have 500 synthetic microbial samples. Then  $k$ -medoids clustering was performed for  $k \in \{1, 2, \dots, 10\}$  using the Jensen-Shannon distance metric (Supplementary Text Sec. 5.1). Silhouette analysis was performed to determine the optimal number of clusters and the clustering results were illustrated in the 2-dimensional principle coordinates plot. For Supplementary Figure S1 the same steps as for the preparation of Figure 1 were performed, but with  $\mathbf{G}$  representing the adjacency matrix of an Erdős-Rényi digraph with mean degree of 20 (mean in-degree of 10 and mean out-degree of 10) and  $s = 1/\sqrt{10}$ . Details on the construction of an Erdős-Rényi digraph can be found in Supplemental Information Section 3.2.1. For Supplementary Figure S2 the same steps as above were performed in Figure 1 but with  $p = 5,000$  local communities.

Figure 3 is a macroscopic analysis of how network structure plays a role in the steady state shift with values of  $\alpha \in (1, 5]$ . For each topology ten different universal matrices  $\mathbf{A}$  were generated. Figure 3(a) shows the results of a complete graph and for each of the ten universal  $\mathbf{A}$  the same steps as in the preparation of Figure 1 were carried out. Figure 3(b) shows the result of an Erdős-Rényi random digraph topology and for each of the ten  $\mathbf{A}$  matrices the same steps as in the preparation of Supplementary Figure S2 were carried out. Figure 3(c) shows results for networks with a power-law out-degree distribution with a mean out-degree of 10, where the out-degree sequence uses the same  $\bar{h}$  in the construction of  $\mathbf{H}$ . More information on the construction of  $\mathbf{G}$  for a power-law out-degree network can be found in Supplementary Text Sec. 3.2.2. Figure 3(d) shows results for networks with a power-law out-degree distribution with mean out-degree of 10 and there is no interaction strength heterogeneity, i.e.  $\mathbf{H}$  is the identity matrix. For this study the Silhouette Index was constructed from normalized steady state data using the Jensen-Shannon distance. Supplementary Figure S9 is the same as Figure 3, but instead of the Silhouette Index, the variance ratio criterion is used with the Jensen-Shannon distance, from normalized steady state abundance (Supplementary Text Sec. 5.4). In Supplementary Figure S10 the Silhouette Index is determined from the Euclidean distance with normalized steady state abundance. Finally, in Supplementary Figure S11 the Silhouette Index is determined by the Euclidean norm with the absolute steady state abundance.

Figure 4 contains a PCoA analysis of the results from Figure 1, but with the Euclidean distance being used instead of the Jensen-Shannon distance, making PCoA equivalent to principle component analysis. This enables us to project the open-loop control trajectories into the principle coordinates (Supplementary Text Sec 5.6). This procedure was also used in the preparation of Supplementary Figure S12.

Supplementary Figures S13 to S15 contain system identification analyses for temporal gut microbiome data of two subjects [13]. The data is publicly available from the metagenomics analysis server MG-RAST:4457768.3-4459735.3 and can also be accessed (as we did) from Qiita (<http://qiita.ucsd.edu>) under study ID 550. The processed data was downloaded as biom file “67\_otu\_table.biom” (2014-11-17 13:18:50.591389). The *Operational Taxonomic Units* (OTUs) were then grouped from the genus level and up, depending on the availability of known classifications for OTUs, and converted to a txt file using MacQIIME version 1.9.0-20140227 with the command `summarize_taxa.py` with the options `-L 6 -a true`. Data was collected over 445 days with 336 fecal samples from Subject A and 131 fecal samples from Subject B. Details on the system identification algorithm are



now given. The dynamics in (2) can be approximated in discrete time as [86]

$$e_i(k) + \log(x_i(t_{k+1})) - \log(x_i(t_k)) = r_i + \sum_{j=1}^n a_{ij}x_j(t_k) \quad (\text{M1})$$

for  $i = 1, 2, \dots, n$  where  $k = 1, 2, \dots, N - 1$  is the sample index,  $N$  is the total number of samples,  $t_k$  is the time stamp of sample  $k$ , and  $e$  is an error term that arises because of the assumption that  $x(t)$  is constant over each interval  $t \in [t_k, t_{k+1})$ . Equation (M1) can be rewritten in terms of a regressor vector

$$\phi(k) = [1, x_1(t_k), x_2(t_k), \dots, x_n(t_k)]^\top,$$

the parameter vector  $\theta_i = [r_i, a_{i1}, a_{i2}, \dots, a_{in}]$  and the log difference  $y_i(k) = \log(x_i(t_{k+1})) - \log(x_i(t_k))$  as

$$e_i(k) + y_i(k) = \theta_i \phi(k).$$

The identification problem can then be defined as finding the parameter matrix estimate  $\hat{\Theta} = [\hat{\theta}_1^\top, \hat{\theta}_2^\top, \dots, \hat{\theta}_n^\top]^\top$  of the true parameter matrix  $\Theta = [\theta_1^\top, \theta_2^\top, \dots, \theta_n^\top]^\top$ . Letting

$$y(k) = [y_1(k), y_2(k), \dots, y_n(k)]^\top$$

be the log difference vector for all species and  $Y = [y(1), y(2), \dots, y(N - 1)]$  be the log difference matrix the system identification problem can be compactly presented as

$$\min_{\hat{\Theta}} \|Y - \hat{\Theta}\Phi\|_F^2 + \lambda \|\hat{\Theta}\|_F^2$$

where  $\Phi = [\phi(1), \phi(2), \dots, \phi(N - 1)]$  is the regressor matrix,  $\|\cdot\|_F$  denotes the Frobenius norm,  $\lambda \geq 0$  is the Tikhonov regularization term [96]. The minimal solution to the above problem can be given directly as

$$\arg \min_{\hat{\Theta}} \left( \|Y - \hat{\Theta}\Phi\|_F^2 + \lambda \|\hat{\Theta}\|_F^2 \right) = F Y^\top (Y Y^\top + \lambda I)^{-1}$$

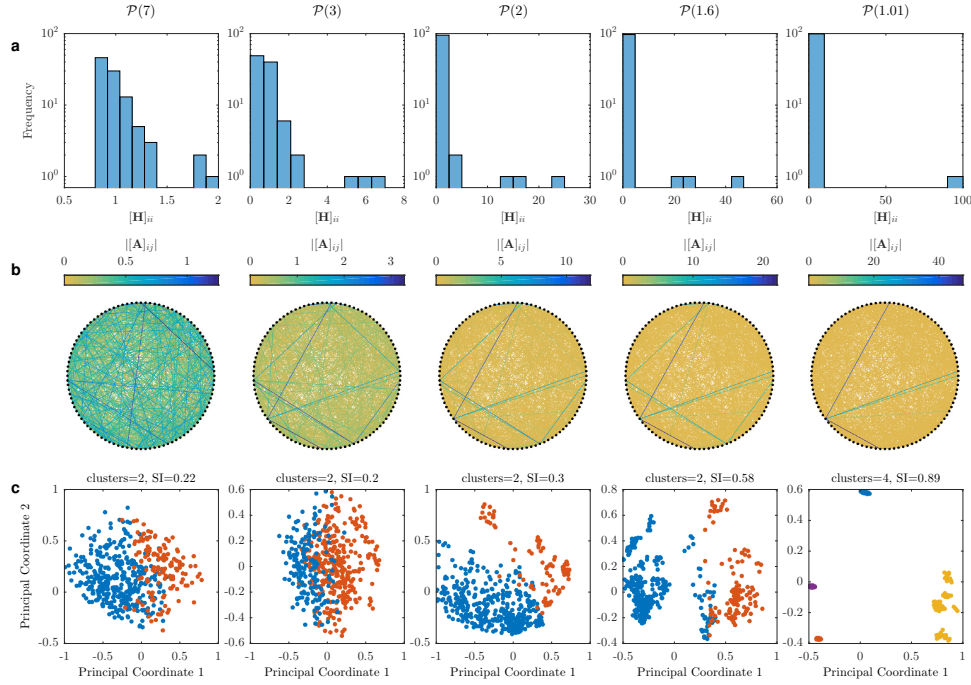
where  $I$  is the identity matrix.

Next we discuss how missing data, zero reads, and  $\lambda$  were chosen. The difference equation in (M1) only uses sample data over two consecutive time samples. Therefore, in the construction of  $Y$  and  $\Phi$  we only include samples that for which there is data from the next day as well. Also, given that the logarithms are used, when a sample has zero reads for a given taxa, a read value of one is inserted. Then relative abundances are computed before the logarithm is taken. Finally we discuss how the regularization parameter is chosen. For Supplementary Figures S13 and S14 the following cross-validation is performed. For Subjects A and B two-thirds of data was used for training and one-third for testing. More precisely, for each  $\lambda$  two-thirds of the data from Subject A and two-thirds of the data from Subject B were used to identify their corresponding dynamical constants. Then the combined error from the two test sets was used to find the optimal  $\lambda$ . The regularization value used in Supplementary Figure S15 is simply the same regularization value used in Supplementary Figure S13.

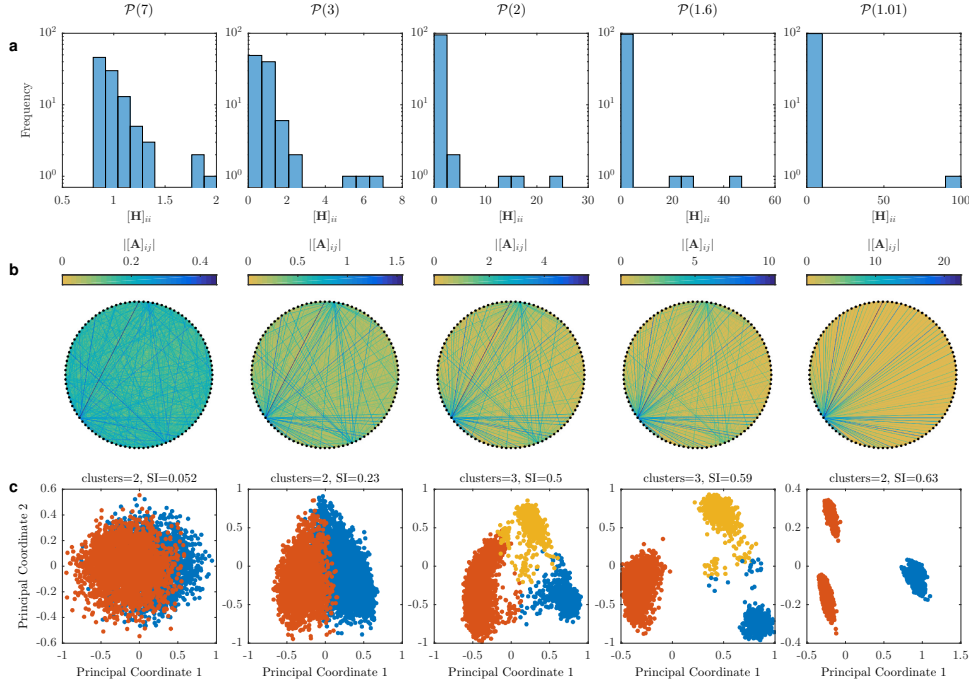


## Chapter 4

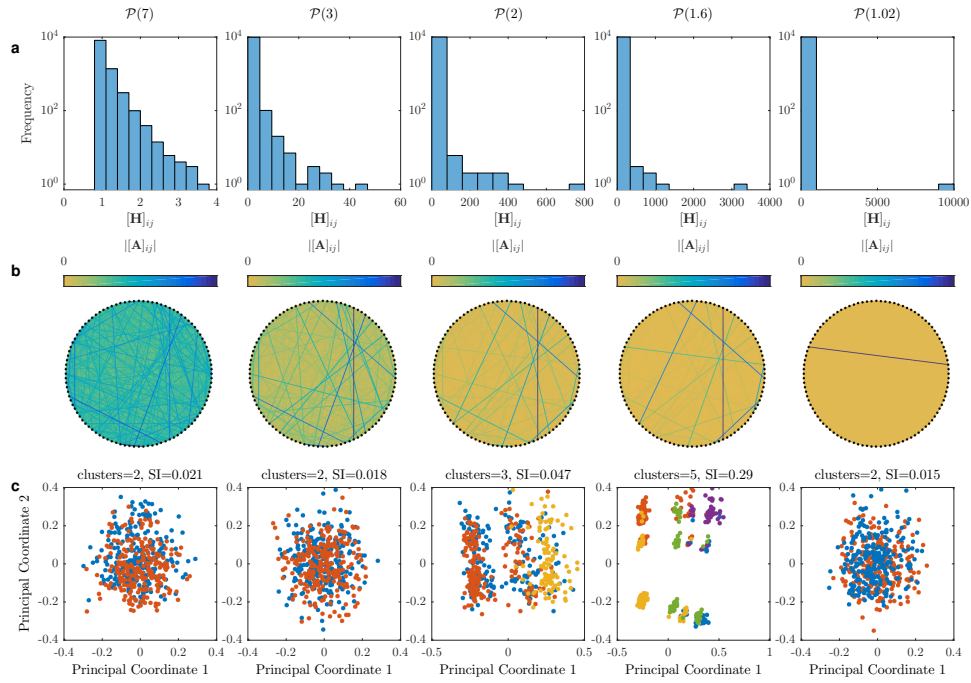
### Supplementary Figures



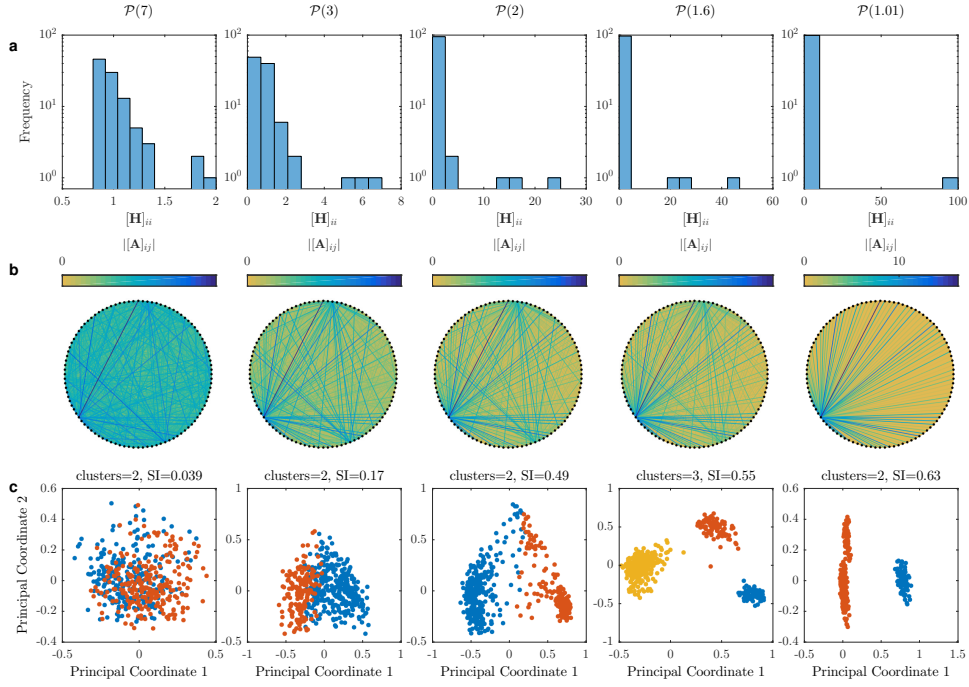
**Supplementary Figure S1:** *Impact of interaction strength heterogeneity on the distinctness of community types.* Same as Figure 1 but with the topology component  $\mathbf{G}$  chosen to be an Erdős-Rényi digraph with a link probability of 0.1 and the scaling factor was set at  $s = 1/\sqrt{10}$ .



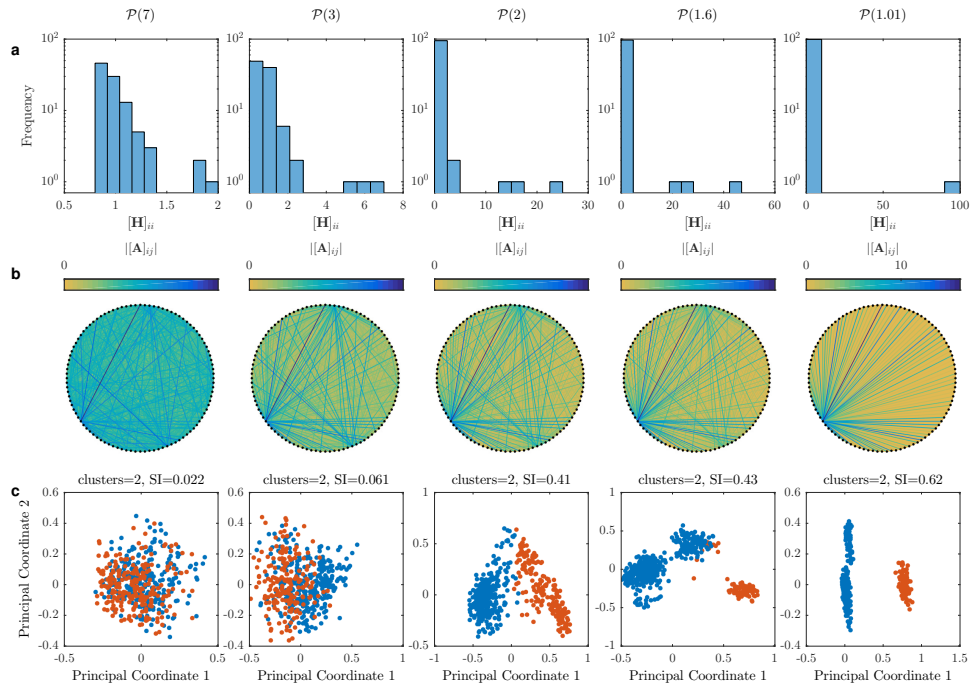
**Supplementary Figure S2:** *Impact of interaction strength heterogeneity on the distinctness of community types.* Same as Figure 1 but with  $p = 5,000$  local communities. Note that it is rather counter-intuitive that for  $\alpha = 1.01$  the Silhouette Index suggests that there are two clusters, while PCoA suggests three clusters. We emphasize that as a typical ordination method, the PCoA just produces a spatial representation of the entities in the dataset, rather than the actual determination of cluster membership [44, 67]. Note that as compared to Figure 1, because there are more samples in this figure, the distinctness of the clusters when  $\alpha = 2$  has shifted to more of a continuous gradient as apposed to distinct clusters.



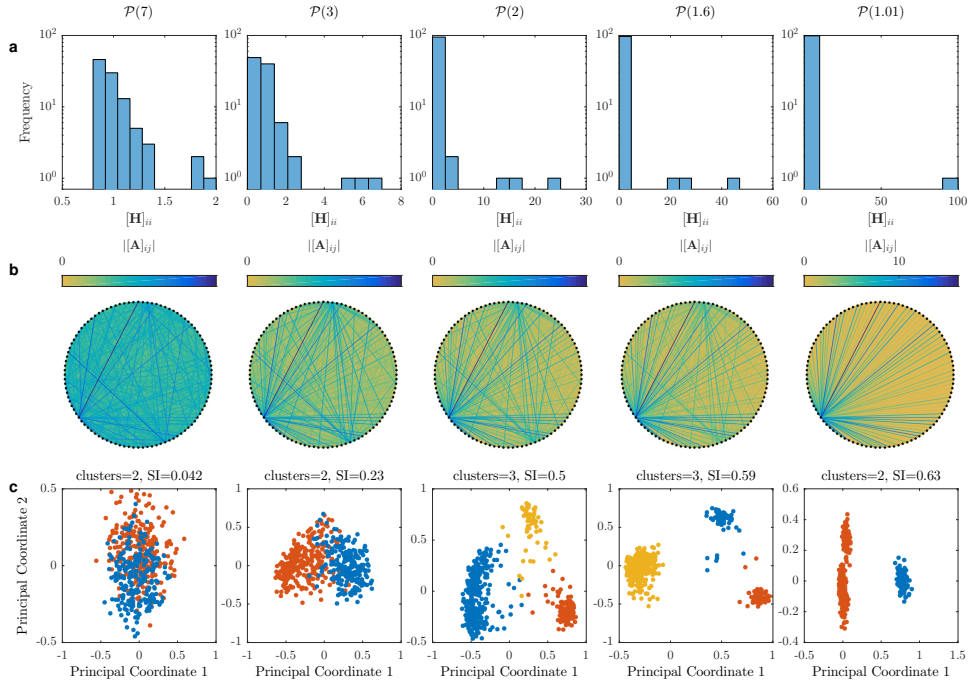
**Supplementary Figure S3:** *Impact of interaction heterogeneity disbursed randomly throughout the network.* The set up is the same as that of Figure 1 but instead of  $\mathbf{H}$  being a diagonal matrix, it is a full matrix, so that individual interactions are scaled randomly from a power-law distribution.



**Supplementary Figure S4: Impact of low levels of migration.** Same as Figure 1 but with a new term  $\lambda(t) \in [0, 1]^n$  added to the dynamics so that now  $\dot{x} = \lambda + \text{diag}(x)(r + Ax)$ . In this example  $\lambda_i \sim \mathcal{U}(0, 0.1)$ . The disturbance is sampled every 0.01 seconds and held constant until the next sample is taken..

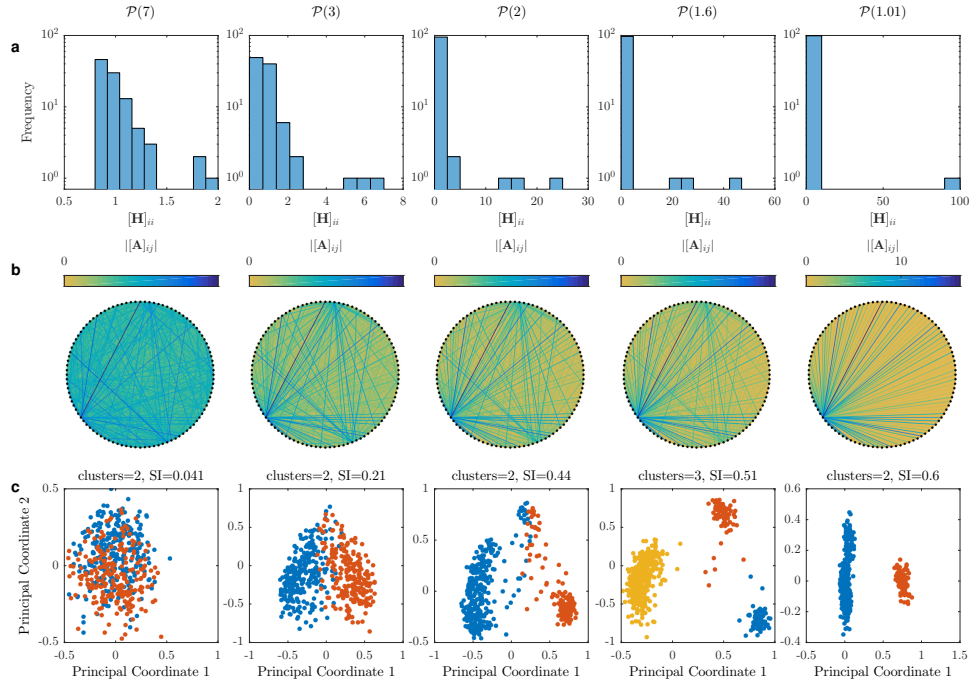


**Supplementary Figure S5: Impact of moderate levels of migration.** Same as Figure 1 but with a new term  $\lambda(t) \in [0, 1]^n$  added to the dynamics so that now  $\dot{x} = \lambda(t) + \text{diag}(x)(r + Ax)$ . In this example  $\lambda_i \sim \mathcal{U}(0, 1)$ . The disturbance is sampled every 0.01 seconds and held constant until the next sample is taken.

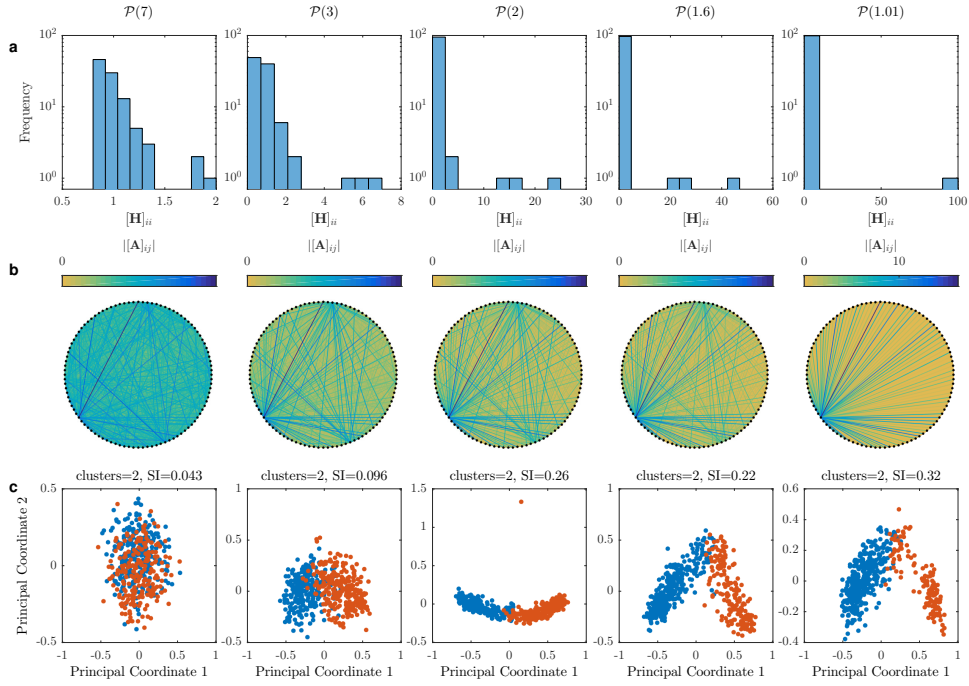


**Supplementary Figure S6:** *Impact of small stochastic disturbance.* Same as Figure 1 but with stochastic Itô dynamics  $dx = \text{diag}(x)(r dt + Ax dt + c dw)$  where  $w$  is a  $n$ -dimensional Brownian motion and  $c$  represents the stochastic disturbance strength. Dynamics were simulated with a discrete time step of 0.01 seconds and  $c = 0.1$ .

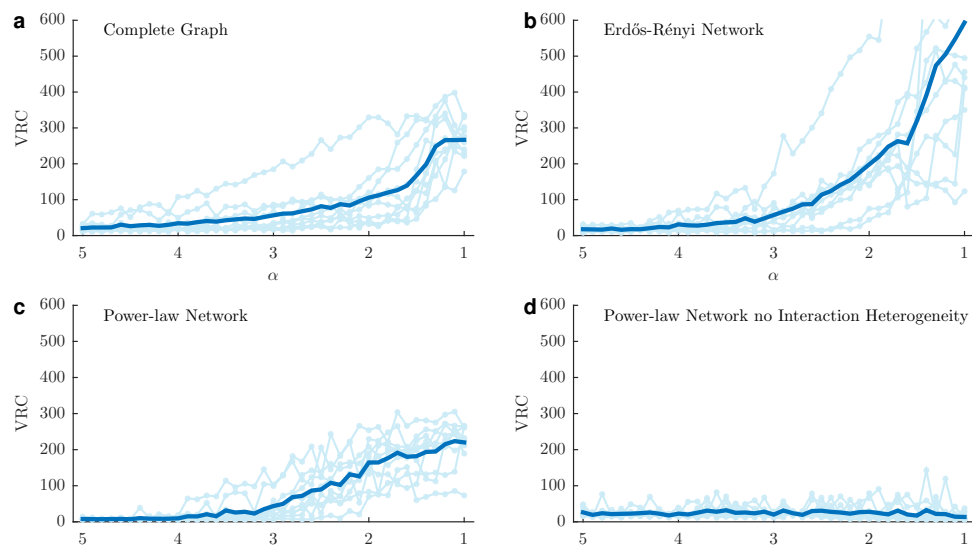




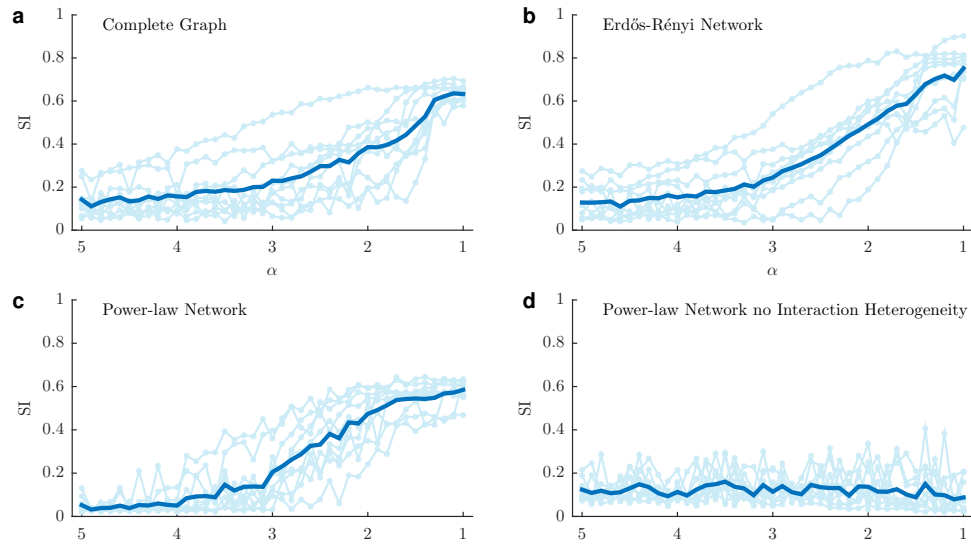
**Supplementary Figure S7: Impact of moderate stochastic disturbance.** Same as Figure 1 but with stochastic Itô dynamics  $dx = \text{diag}(x)(r dt + Ax dt + c dw)$  where  $w$  is a  $n$ -dimensional Brownian motion and  $c$  represents the stochastic disturbance strength. Dynamics were simulated with a discrete time step of 0.01 seconds and  $c = 0.5$ .



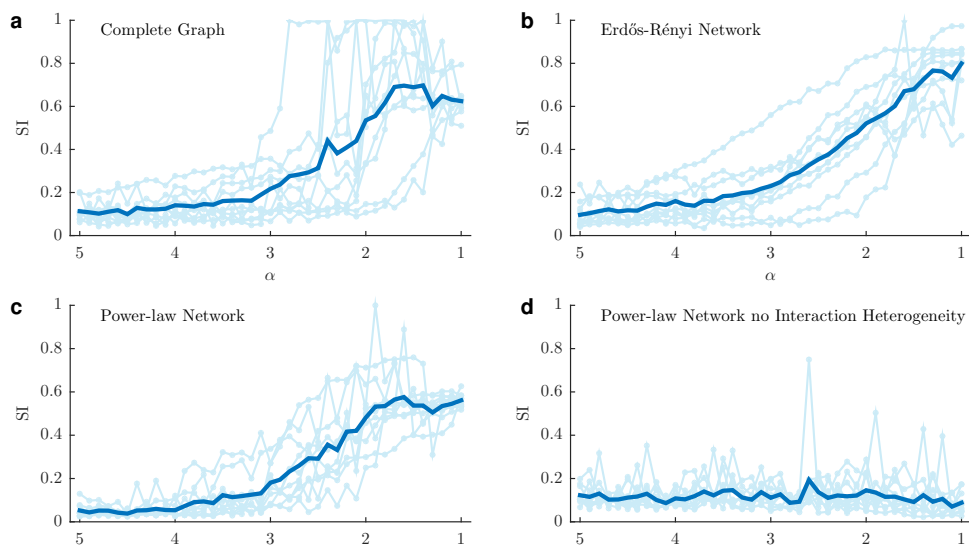
**Supplementary Figure S8: Impact of large stochastic disturbance.** Same as Figure 1 but with stochastic Itô dynamics  $dx = \text{diag}(x)(r dt + Ax dt + c dw)$  where  $w$  is a  $n$ -dimensional Brownian motion and  $c$  represents the stochastic disturbance strength. Dynamics were simulated with a discrete time step of 0.01 seconds and  $c = 1$ .



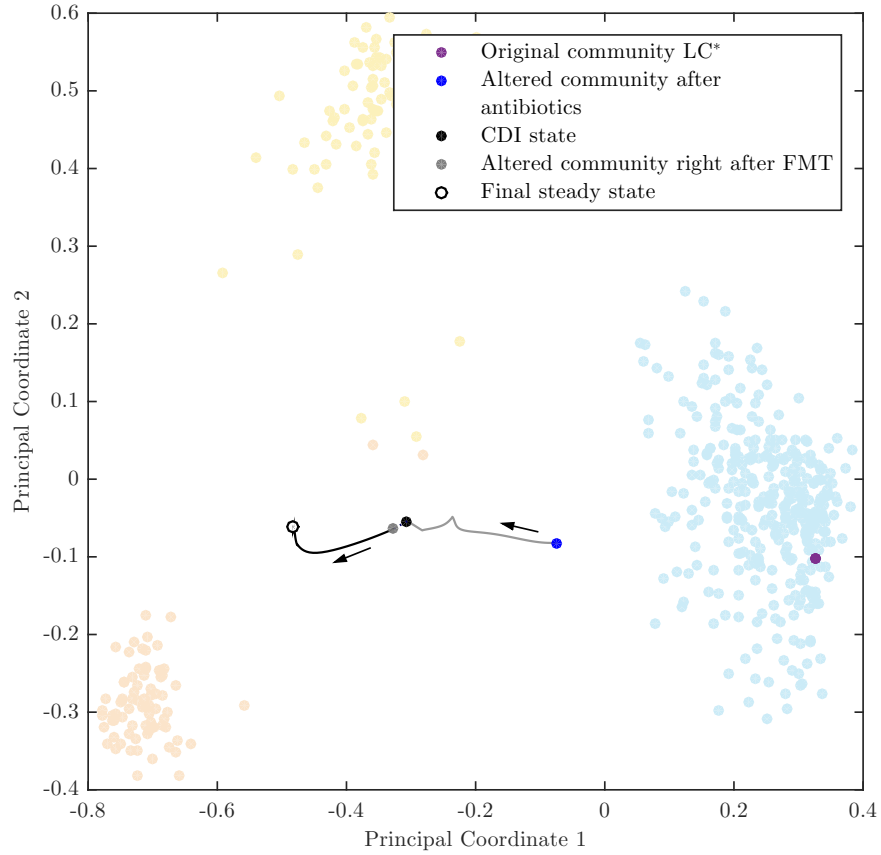
**Supplementary Figure S9:** *Impact of network structure on the distinctness of community types.* The same as Figure 3 with the *Variance Ratio Criterion* (VRC) used as apposed to the Silhouette Index for the clustering measure. See Supplemental Information §5.4 for details on the VRC.



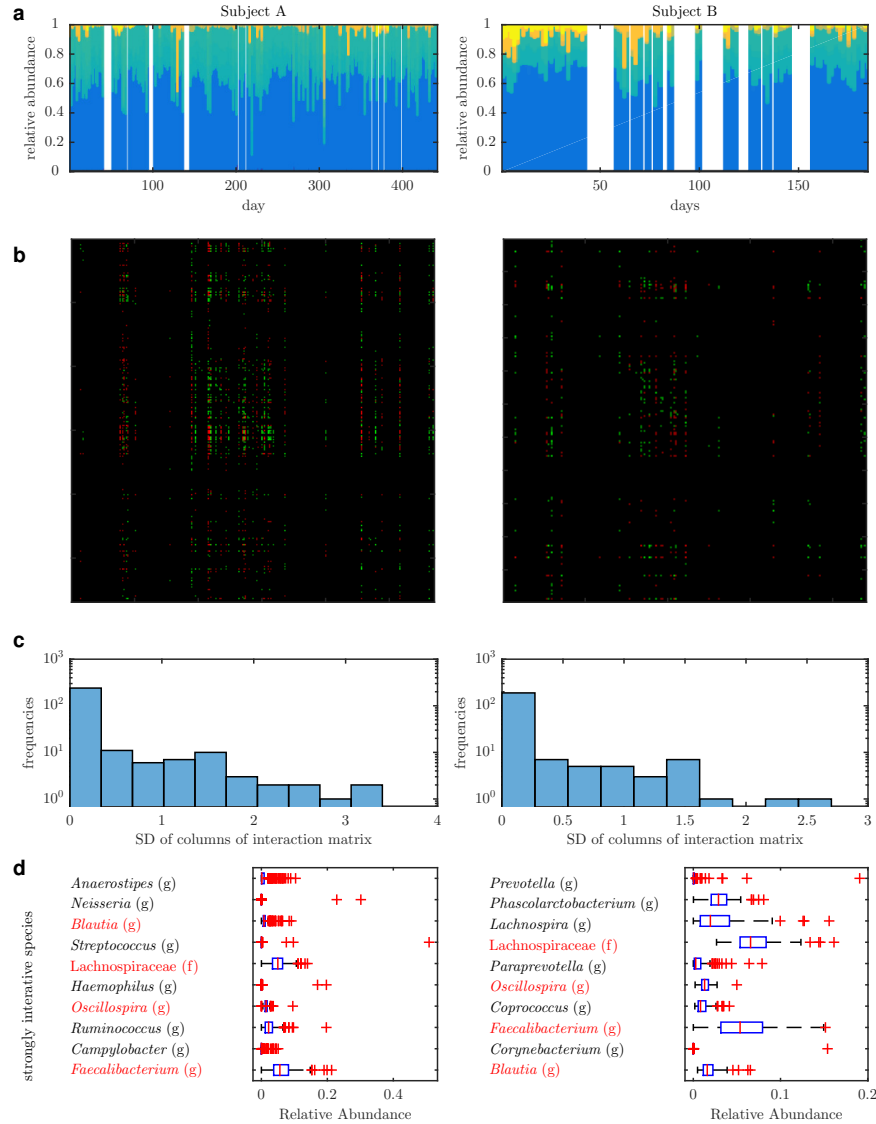
**Supplementary Figure S10:** *Impact of network structure on the distinctness of community types.* The same as Figure 3 with the Euclidean distance metric used instead of the Jensen-Shannon distance metric.



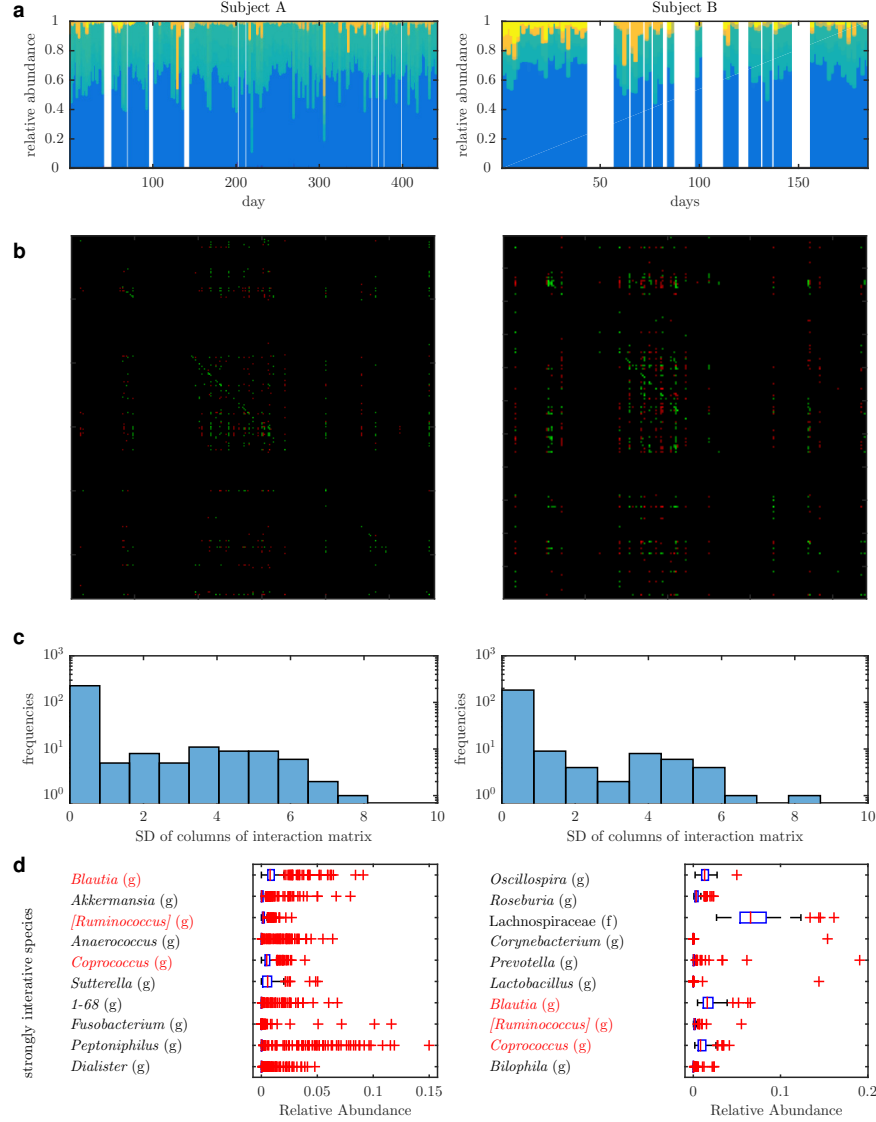
**Supplementary Figure S11:** *Impact of network structure on the distinctness of community types.* The same as Figure 3 with the Euclidean distance metric used instead of the Jensen-Shannon distance metric and absolute abundance used instead of relative abundance.



**Supplementary Figure S12:** *Unsuccessful Fecal Microbiota Transplantation.* Similar to Scenario 3 shown in Figure 4a, but during the FMT, the SISs (60 and 51) of the donor's local community in the orange cluster were not transplanted to the CDI state (black dot). This FMT resulted in a slightly altered community (gray dot) and the system eventually evolved to a steady state (white dot) which is not in the orange cluster. Hence the FMT failed.

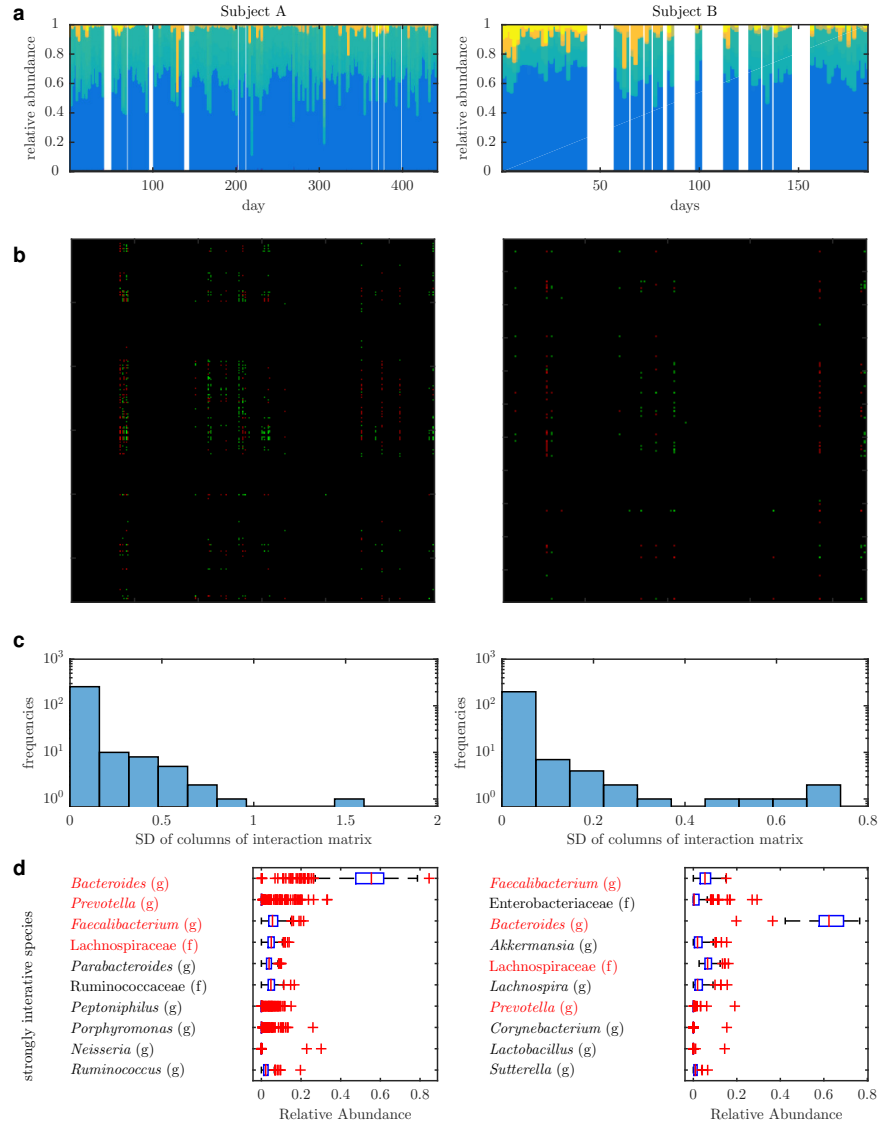


**Supplementary Figure S13: System Identification, Tikhonov Regularization**  
 $\lambda = 0.0423$ . System identification was performed on the stool samples from the longitudinal data in [13] for two subjects as described in the Supplementary Methods where  $\lambda$  was determined by cross-validation. (a) Visualization of microbial taxa in terms of relative abundances versus day sample was taken. (b) Heat map of the interaction matrix for top 100 SISs. (c) Histogram of *Standard Deviation* (SD) of the columns of the interaction matrix. (d) List of top ten SISs in descending interaction strength (defined by the SD of each column in the interaction matrix) with relative abundances over all samples shown as a box plot. The banded structure shown in the heat map supports the assertion that SISs do exist in the gut microbiome. However this banded structure is also seen when the dates of the sample collections are permuted, see Supplementary Figure S14 and S15.



**Supplementary Figure S14: System Identification, Day Swap, Tikhonov Regularization  $\lambda = 0.0057$ .** System identification was performed on the stool samples from the longitudinal data in [13], but with the collection dates permuted,  $\lambda$  was determined by cross-validation on the permuted data. (a) Visualization of microbial taxa in terms of relative abundances versus day sample was taken (not permuted samples). (b) Heat map of the interaction matrix for top 100 SISs. (c) Histogram of *Standard Deviation* (SD) of the columns of the interaction matrix. (d) List of top ten SISs in descending interaction strength (defined by the SD of each column in the interaction matrix) with relative abundances over all samples shown as a box plot. Even though the sample days have been permuted the banded structure still persists.





**Supplementary Figure S15: System Identification, Day Swap, Tikhonov Regularization**  $\lambda = 0.0423$ . System identification was performed on the stool samples from the longitudinal data in [13], but with the collection dates permuted,  $\lambda$  was selected to be the same as in Supplementary Figure S13. (a) Visualization of microbial taxa in terms of relative abundances versus day sample was taken (not permuted samples). (b) Heat map of the interaction matrix for top 100 SISs. (c) Histogram of *Standard Deviation* (SD) of the columns of the interaction matrix. (d) List of top ten SISs in descending interaction strength (defined by the SD of each column in the interaction matrix) with relative abundances over all samples shown as a box plot. For the permuted data when  $\lambda$  is larger than the optimal value from the cross-validation the identification method biases towards making the most abundant species also the SISs.



## Chapter 5

# Supplementary Text

### 1 Introduction

The purpose of this supplement is to give a self-contained, rigorous account of the topics needed to fully understand the main text. The supplement is organized as follows. Section 2 contains some remarks on notation. Section 3 covers the basics of random variables, gives details on how one constructs a well defined random variable from a power-law distribution, and covers some basic results in random matrix theory, Wigner's circle and semi-circle laws. Section 4 contains details on the definition of asymptotic stability, how this relates to the generalized Lotka-Volterra dynamics, some recent results on diagonal stability are discussed, and new results regarding the stochastic stability of Lotka-Volterra dynamics is discussed. Section 5 gives the details of the multi-dimensional scaling analysis and clustering algorithms employed in this study. Section 6 gives the details of our modeling approach. Section 7 contains simulation results that are intended to complement the figures in the main text and the supplementary figures.

Much of the foundational material is well known in each of their respective areas. The diagonal stability result for random matrices has never been discussed in the context of Lotka-Volterra dynamics. The discussions regarding stable steady state shift for Lotka-Volterra dynamics are the first of their kind as well.

Those familiar with probability theory need only read §3.3 for a refresher on Wigner random matrices and the rest of §3 can be skipped. Those familiar with stability need only visit §4.3 and §4.5. That being said, even the seasoned stability theorist may find new things in the stochastic subsection of §4.4. Section 5 can be skipped for those familiar with the commonly used tools in clustering analysis.

### 2 Notation

Throughout we denote the real numbers as  $\mathbb{R} = (-\infty, \infty)$ , and the complex numbers as  $\mathbb{C}$ . A number  $z \in \mathbb{C}$  if  $z = x + iy$  where  $x, y \in \mathbb{R}$  and  $i \triangleq \sqrt{-1}$ . The positive real numbers are denoted as  $\mathbb{R}_{>0} \triangleq (0, \infty)$  and the non-negative reals as  $\mathbb{R}_{\geq 0} \triangleq [0, \infty)$ . The  $n$ -dimensional reals are denoted as  $\mathbb{R}^n$ ,  $\mathbb{R}_{\geq 0}^n$  denotes the  $n$ -dimensional space of positive vectors which we will refer to as the *positive orthant*, and  $\mathbb{R}_{\geq 0}^n$  as the *non-negative orthant*. An  $m \times n$  matrix  $A$  of values in the reals is denoted as  $A \in \mathbb{R}^{m \times n}$ . The element of  $A$  in the  $i$ -th row and the  $j$ -th column will often be denoted with lower case letters as  $a_{ij} = [A]_{ij}$ . If there is an issue with formatting and it is not clear that both elements  $i$  and  $j$  are subscripts in the previous notation then it is equivalent to denote  $a_{ij}$  as  $a_{i,j}$ . The dual notation for an  $n \times n$  matrix constructed from elements  $a_{ij}$  is denoted as  $A = (a_{ij})_{1 \leq i, j \leq n}$ .

The superscript  $(\cdot)^T$  is used to denote transpose. The components of an  $n$  dimensional vector  $x$  are defined as follows  $x = [x_1, x_2, \dots, x_n]^T$ . We will often make use of the following nonstandard subscript notation,  $x_{j:k}$  to denote the vector obtained from taking the  $j$ -th element to the  $k$ -th element of  $x$ .<sup>1</sup> As an example, consider  $y = [1, 3, 5]^T$ , then  $y_{2:3} = [3, 5]^T$ . The Euclidean norm of a vector  $x \in \mathbb{R}^n$  is defined as  $\|x\| \triangleq (\sum_{i=1}^n x_i^2)^{1/2}$ . When applied to a matrix  $A \in \mathbb{R}^{m \times n}$  the norm is an induced norm  $\|A\| \triangleq \sup_{\|x\|=1} \|Ax\|$ .

---

<sup>1</sup>This notation was motivated by the index notation used in Matlab.

For a square symmetric matrix  $P = P^\top \in \mathbb{R}^{n \times n}$  the inequality  $< (\leq)$  is used as  $P < (P \leq 0)$  if and only if  $x^\top P x < 0 (x^\top P x \leq 0)$  for all  $x \in \mathbb{R}^n$ .

The notation for union and intersection of sets is  $\cup$  and  $\cap$  respectively. For a set  $A \subset U$  the complement of  $A$  is defined as  $A^c = \{y \in U \mid y \notin A\}$ . The set minus notation is defined as  $A \setminus B = A \cap B^c$ . Let  $A$  be a set, then  $|A|$  denotes the cardinality of that set. If for instance  $A = \{1, 2, 4\}$ , then  $|A| = 3$ .

Given that  $x$  will primarily be defined as the state variable, we will use  $Y$  when discussing a generic random variable. The probability distribution for  $Y$  is denoted as  $\mu_Y$  for which we will generically denote its probability density function as  $f(y)$  where  $\mu_Y = \int f(y) dy$ . When given a probability density function  $f(y)$ , the notation  $Y \sim f(y)$  reads as the random variable  $Y$  is drawn from the probability density function  $f$ . Also, if we generically denote the standard normal distribution as  $\mathcal{N}(0, 1)$ , then with a slight abuse of notation we can write  $Y \sim \mathcal{N}(0, 1)$  to denote that  $Y$  is drawn from the standard normal distribution. The notation  $X \equiv Y$  is used to denote when two random variables are drawn from the same distribution.

We will also make use of the following asymptotic notation.  $f(n) = O(g(n))$  if there exists a  $C$  independent of  $n$  such that  $\|f(n)\| \leq C\|g(n)\|$  for  $n$  sufficiently large. Another form of asymptotic notation that will be borrowed is the following,  $f(n) = o(g(n))$  if there exists a  $c(n) \geq 0$  where  $\lim_{n \rightarrow \infty} c(n) = 0$  and  $\|f(n)\| \leq c(n)\|g(n)\|$ .

Results from dynamics and control, to probability and random matrix theory will be called upon in this work. If the notation of a particular section seems to overlap, it is assumed that the convention from the field of origin overrides. For instance, when discussing dynamics a capital letter will denote a matrix, however capital letters are used in probability theory when denoting generic random variables.

### 3 Random Variables and Random Matrices

#### 3.1 Primer on Random Variables

We will use the following three distributions in constructing the random interaction matrices for the microbial communities: the uniform distribution taking values in the interval  $[0, 1]$ , the normal (Gaussian) distribution of mean 0 and variance  $\sigma^2$  and a power-law distribution with minimum value 1 and exponent  $-\alpha$ . One can define a random variable just with a variable space and ignore the underlying sample space [92, §1.1.2]. We however introduce the full probabilistic machinery into the discussion. Formality in this section is two fold. First, to give the reader confidence that the random variables generated from the power-law distribution are indeed well defined random variables [15]. Second, the Wigner circle law and semi-circle law can not be defined without this machinery, and these two laws are fundamental to understanding stability results that are presented in the next section. The following is a nearly verbatim presentation of probability spaces following [91, 92].

Let  $\Omega$  be a sample space. When given a measure and a  $\sigma$ -algebra,  $\Omega$  becomes a probability space  $\Omega = (\Omega, \mathcal{F}, \mathbf{P})$ , where  $\mathcal{F}$  is a  $\sigma$ -algebra of subsets of  $\Omega$  and  $\mathbf{P}$  is a probability measure. The probability measure along with the sample space satisfy the following equality  $\mathbf{P}(\Omega) = 1$ . Events  $E$  are then taken from the  $\sigma$ -algebra and will have a probability of occurring, i.e.  $E \in \mathcal{F}$  and  $E \mapsto \mathbf{P}(E)$  where  $\mathbf{P}(E) \in [0, 1]$ . A random variable  $Y$  takes values in a measurable space  $R = (R, \mathcal{R})$  where  $\mathcal{R}$  is a  $\sigma$ -algebra of subsets of  $R$ . We will also refer to  $R$  as the variable space. The formal definition of a random variable is a map  $Y : \Omega \rightarrow R$  where  $Y$  is measurable. When one asks for the probability that event  $Y$  is in  $S \in \mathcal{R}$ , we are interested in the probability of event  $E = Y^{-1}(S)$  occurring. From the definition of our probability space  $\Omega$ , this is simply  $\mathbf{P}(Y^{-1}(S))$ . Note that this is equivalent to  $\mathbf{P}(\{\omega \in \Omega : Y(\omega) \in S\})$  which we can unambiguously denote in shorthand as  $\mathbf{P}(Y \in S)$  [92, §1.1.2]. This notation makes no reference to the original sample space and thus can be used without ambiguity when discussing the probability of an event occurring. All of our probability spaces and variable spaces will be subsets of  $\mathbb{R}$  or  $\mathbb{C}$ . Thus we will implicitly use the Borel  $\sigma$ -algebra. Consequently, the specific  $\sigma$ -algebra of interest will no longer be denoted.

We now give rigorous definitions for the probability measure  $\mu_Y$  of a random variable  $Y \in R = (R, \mathcal{R})$  and the corresponding probability density function  $f$  associated with  $\mu_Y$ .

The distribution of  $Y$  is defined as

$$\mu_Y(S) \triangleq \mathbf{P}(Y \in S).$$

Given the above definition we define the probability density function as

$$\mu_Y(S) = \int_S f(y) dy.$$

The cumulative distribution function is defined as

$$F_Y(y) \triangleq \mathbf{P}(Y \leq y) = \mu_Y((-\infty, y]).$$

The expected value of a random variable  $Y \sim f(y)$  (taking values from the probability density function  $f$ ) is defined as

$$\begin{aligned} \mathbf{E}Y &\triangleq \int_{\mathcal{R}} y d_Y(y) \\ &= \int_{\mathcal{R}} y f(y) dy \\ &= \int_{\mathcal{R}} \mathbf{P}(Y \geq \lambda) d\lambda. \end{aligned} \tag{T1}$$

Sometimes parentheses are used, i.e.  $\mathbf{E}(Y)$ , for clarity. The variance of a distribution is defined as

$$\mathbf{Var}(Y) \triangleq \mathbf{E}|Y - \mathbf{E}(Y)|^2. \tag{T2}$$

**Definition 1** ([Almost Surely]). An event  $E$  occurs *almost surely* if  $\mathbf{P}(E) = 1$ .

### 3.1.1 Distributions of interest

The uniform distribution generates random variables on the measure space  $\mathcal{R} = [0, 1]$  with the following probability density function

$$u(y) = \begin{cases} 1 & \text{if } y \in [0, 1] \\ 0 & \text{otherwise} \end{cases}.$$

A pseudorandom variable in  $[0, 1]$  can be generated uniformly using the MATLAB command `random('unif', 0, 1)`. We will use the shorthand notation  $\mathcal{U}(0, 1)$  to denote a uniform distribution taking values in  $[0, 1]$ .

The normal distribution with mean 0 and standard deviation  $\sigma$  generates random variables on the measure space  $\mathbb{R}$  and satisfies the well known probability density function

$$n(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}.$$

A pseudorandom variable with mean 0 and variance  $\sigma^2$  can be generated using the MATLAB command `random('norm', 0, sigma)`. We will use the shorthand notation  $\mathcal{N}(0, \sigma^2)$  to denote a normal distribution with mean 0 and standard deviation  $\sigma$ . A log-normal distributed random variable is simply one in which  $Y = e^X$  where  $X \sim \mathcal{N}(0, \sigma^2)$ .

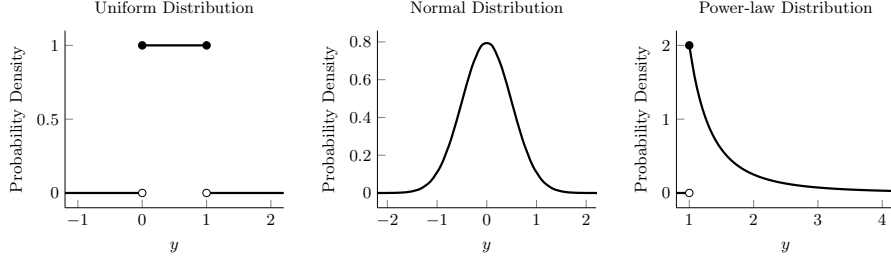
The generic power-law used in this work generates random variables on the measure space  $[1, \infty)$  with the following probability density function

$$p(y) = (\alpha - 1)y^{-\alpha} \tag{T3}$$

where  $\alpha > 1$  [15, Equation (2.2)]. We will use the following short hand notation  $\mathcal{P}(\alpha)$  to denote a power-law distribution with exponent  $-\alpha$ .

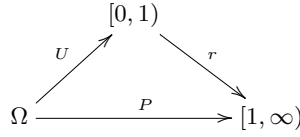
Later we will generate power-law distributions from uniform distributions, and thus we have the following simple result. Let  $U$  be a uniform random variable from the set  $[0, 1]$ , then we can generate a random variable  $P$  with a power-law distribution in  $[1, \infty)$  using the following measurable monotonic function  $r : [0, 1] \rightarrow [1, \infty)$ ,

$$r(U) = (1 - U)^{\frac{1}{1-\alpha}}. \tag{T4}$$



**Supplementary Text Figure T1:** Illustration of the distributions for the three types of probability functions used in this work, from left to right: uniform, normal, and power-law.

Now, letting  $P \triangleq r(U)$  we can see that  $P$  indeed satisfies all of the requirements to be a well defined random variable.  $P$  is a measurable function defined from a probability space (the variable space of  $U$  with a probability measure  $\mathbf{P}$ ) to an event space by a measurable function  $r$ . This mapping for the generation of a random variable  $P$  with powerlaw distribution is illustrated with the following commutative diagram.



Random variables satisfying a power-law probability density function can be generated in MATLAB using the following line `(1-random('unif',0,1))^(1/(1-alpha))`.

The nice trick in (T4) is introduced in the literature without proof, and is simply denoted as following from the fundamental transformation law of probabilities [77, §7.3]. Given the rigorous introduction of probability theory at the beginning of this section, we can however directly derive this result by analyzing the probability that  $P > y$  under the assumption that there exists a monotonically increasing bijection  $r : [0, 1) \rightarrow [1, \infty)$ , which yields

$$\begin{aligned}
 \int_y^\infty p(x)dx &= \mathbf{P}(P > y) \\
 &= \mathbf{P}(r(U) > y) \\
 &= \mathbf{P}(U > r^{-1}(y)) \\
 &= \int_{r^{-1}(y)}^\infty u(x)dx.
 \end{aligned} \tag{T5}$$

The transition from line 2 to line 3 in the above equality follows from the fact that  $r$  is monotonically increasing and thus when the inverse is taken the sign of the inequality is preserved. Integrating the above equality it follows that

$$y^{1-\alpha} = 1 - r^{-1}(y).$$

Substituting  $z = r^{-1}(y)$  and solving for  $y$  in the above equality it follows that

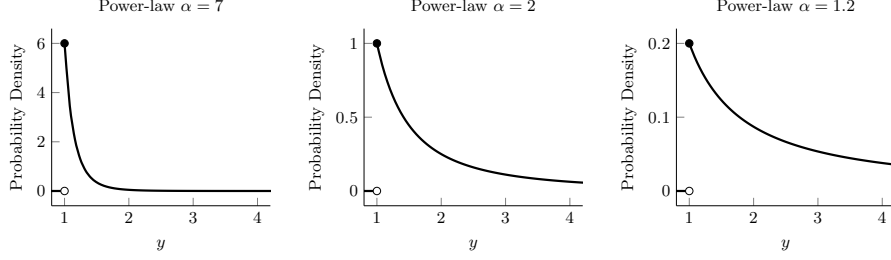
$$y = (1 - z)^{\frac{1}{1-\alpha}}$$

which proves the relation in (T4).

### 3.2 Random Network Models

Formally a digraph  $\mathcal{G}$  is defined by the double  $(\mathcal{V}, \mathcal{E})$  where  $\mathcal{V} = \{1, 2, \dots, n\}$  is the vertex set and the directed edges are defined by the ordered pairs  $(i, j) \in \mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ . An element  $(i, j) \in \mathcal{E}$  if and only if there is a directed edge from vertex  $i$  to vertex  $j$ . We are primarily interested in the adjacency matrix  $\mathcal{A}$  of the digraph which is defined as

$$[\mathcal{A}]_{ij} = \begin{cases} 1 & \text{if } (j, i) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}.$$



**Supplementary Text Figure T2:** Illustration of the power-law distribution from low heterogeneity to high heterogeneity.

When  $[\mathcal{A}]_{ij} = 1$  for all  $i$  and  $j$  the digraph is said to be *complete*. Note that we allow for self-loops in our construction. Two other digraph topologies to be discussed shortly are the Erdős-Rényi (Gilbert) random digraph and the power-law degree digraph.

### 3.2.1 Erdős-Rényi (Gilbert) Digraph

An Erdős-Rényi (Gilbert) digraph is a digraph  $\mathcal{G}(n, p)$  of  $n$  nodes, where the probability of a directed edge from node  $i$  to node  $j$  is  $p$  for any  $i, j \in \mathcal{V}$ . The adjacency matrix for this model is constructed as follows. Let  $G \in [0, 1]^{n \times n}$  be a matrix with elements independently sampled from the uniform distribution between 0 and 1,  $[G]_{ij} \sim \mathcal{U}(0, 1)$ . Then let  $\mathcal{A}$  be defined as follows

$$[\mathcal{A}]_{ij} = \begin{cases} 1 & \text{if } [G]_{ij} < p \\ 0 & \text{otherwise} \end{cases}.$$

If one is interested in defining an Erdős-Rényi model for a 100 node digraph with an expected mean in-degree (or out-degree) of 10, then simply set  $p = 10/100$  in this construction.

While the above model is often credited to Erdős and Rényi [28, 29], it was actually first presented by Gilbert in [36]. We will follow convention however and simply refer to this random network model as the *Erdős-Rényi* (ER) model. The adjacency matrix for this model can be generated in MATLAB using the following boolean expression `rand(n,n) < p`.

### 3.2.2 Power-law Out-degree Digraph

In this section we outline how one can generate the adjacency matrix for a power-law out-degree digraph. Let  $h = [h_1, h_2, \dots, h_n]^T$  be the column vector of out-degrees for nodes  $\{1, 2, \dots, n\}$ . If one is interested in having a digraph with a power-law out-degree of exponent  $-\alpha$  with the mean out-degree approximately  $d$ , then setting

$$h_i = \min \left\{ \left\lceil d \frac{[\bar{h}]_i}{\text{mean}(\bar{h})} \right\rceil, n \right\} \quad (\text{T6})$$

is sufficient, where  $[\bar{h}]_i \sim \mathcal{P}(\alpha)$ ,  $i = 1, 2, \dots, n$ . Note that  $\lceil \cdot \rceil$  is the ceiling operator. We also note that this will not guarantee that the mean out-degree is  $d$  for any finite sized digraph. Finally the adjacency matrix is constructed by selecting  $h_i$  random elements in column  $i$  of  $\mathcal{A}$  and setting them to 1.

This method of generating power-law degree distributions leaves much to be desired. It is not based on any known theory for power-law degree graphs [9, 58]. So as to be able to compare Erdős-Rényi digraphs with the above power-law digraphs we have normalized by degree, which is not a common practice in the literature either. We note also that few authors have rigorously analyzed power-law digraphs with the exception of [9, §11] and [8].

## 3.3 Spectrum of Random Matrices

Two classic results from Wigner [99–101] are now discussed. First we define the *Empirical Spectral Distribution* (ESD) of an  $n \times n$  real valued matrix  $A$  as  $\mu_A : \mathbb{C} \rightarrow \mathbb{N}$

$$\mu_A(z) \triangleq \frac{1}{n} |\{1 \leq i \leq n : \text{Re } \lambda_i(A) \leq \text{Re } z, \text{Im } \lambda_i(A) \leq \text{Im } z\}|.$$

Recall that when applied to a finite set  $|\cdot|$  denotes the cardinality of that set. The ESD simply counts the number of eigenvalues of  $A$  within radius  $|z|$  of the origin. A weaker version of [93, Theorem 1.10] is now stated.

**Theorem 1.** *Let  $B_n$  be a real  $n \times n$  matrix whose elements are independently and identically distributed random variables with mean 0 and variance 1. Then it follows that  $\mu_{B_n/\sqrt{n}}$  converges almost surely to the uniform disk in the complex plane,  $\mathbf{1}_{|z| \leq 1}$ , with probability 1. Let  $R$  denote the variable space for the i.i.d. random variables, then*

$$\mathbf{P} \left( \limsup_{n \rightarrow \infty} \left| \int_{\mathbb{C}} h(z) d\mu_{B_n/\sqrt{n}}(z) - \int_{\mathbb{C}} h(z) \mathbf{1}_{|z| \leq 1} dz \right| \leq \epsilon \right) = 1$$

for all  $\epsilon > 0$  and every bounded continuous function  $h : R \rightarrow \mathbb{C}$ .

The semi-circle distribution is defined as  $\mu_{sc}(S) \triangleq \int_S \rho_{sc}(y) dy$  where the probability density function is defined as

$$\rho_{sc}(y) = \begin{cases} \frac{1}{2\pi} (4 - y^2)_+^{1/2}, & |y| \leq 2 \\ 0, & |y| > 2 \end{cases}$$

We now state a more conservative version of [90, Theorem 5].

**Theorem 2.** *Let  $M_n$  be a symmetric  $n \times n$  matrix whose diagonal and upper right elements are independently and identically distributed random variables with mean 0 and variance 1. Then it follows that  $\mu_{M_n/\sqrt{n}}$  converges in the sense of probability to  $\mu_{sc}$ . Let  $R$  denote the variable space for the i.i.d. random variables, then*

$$\liminf_{n \rightarrow \infty} \mathbf{P} \left( \left| \int_{-2}^x h(y) d\mu_{M_n/\sqrt{n}}(y) - \int_{-2}^x h(y) d\mu_{sc}(y) \right| \leq \epsilon \right) = 1$$

for all  $\epsilon > 0$  and every bounded continuous function  $h : R \rightarrow \mathbb{R}$ .

A conservative version of [5, Theorem A] is now stated

**Theorem 3.** *Let  $M_n$  be a symmetric  $n \times n$  matrix whose upper right elements are independently and identically distributed random variables with mean 0 and variance  $\sigma^2$ , and whose diagonal elements are independently and identically distributed with mean 0 and finite variance. Then it follows that  $\sup_{1 \leq i \leq n} |\lambda_i(M_n)| = 2\sigma\sqrt{n}(1 + o(1))$  asymptotically almost surely. Stated another way,*

$$\liminf_{n \rightarrow \infty} \mathbf{P} \left( \sup_{1 \leq i \leq n} |\lambda_i(M_n)| = 2\sigma\sqrt{n}(1 + o(1)) \right) = 1.$$

Theorem 1 states that the spectral distribution of a random matrix with elements drawn from a normal distribution with mean 0 and variance  $1/n$  converges to the unit disk centered at the origin of the complex plain. For the same random matrix but with the added assumption of symmetry, the spectrum converges to the line segment  $[-2, 2]$  on the real line, see Figure T3. The spreading of the spectrum from a diameter of 2 to a diameter of 4 can be explained by the fact that all 2-cycles in the matrix now have a positive loop game. Positive feedback for 2-cycles always repels the eigenvalues along the real axis [30].

## 4 Stability

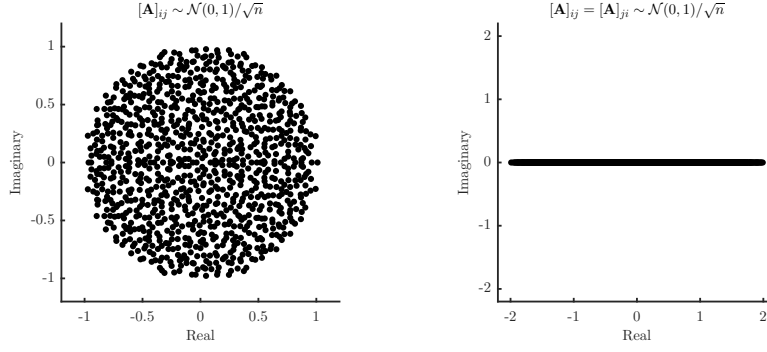
### 4.1 Primer on Stability

Consider the time-varying dynamical system defined by

$$\begin{aligned} \dot{x}(t) &= f(x(t), t) \\ x(t_0) &= x_0 \end{aligned} \tag{T7}$$

where  $x : \mathbb{R} \rightarrow \mathbb{R}^n$  is the state vector,  $t$  is time,  $t_0$  is the initial time, and  $\dot{(\cdot)} \triangleq \frac{d}{dt}(\cdot)$ . Let  $x^* \in \mathbb{R}^n$  be the equilibrium solution so that  $f(x^*, t) = 0$  for all  $t$ . The solution to the ordinary differential equation in (T7) is a transition function  $\phi(t; x_0, t_0)$  such that  $\phi(t_0; x_0, t_0) = x_0$  and  $\dot{\phi}(t; x_0, t_0) = f(\phi(t; x_0, t_0), t)$ . For existence and uniqueness conditions see [16]. Below we give the various definitions of stability as defined in [40, 48, 63, 70].





**Supplementary Text Figure T3:** Eigenvalues for (left) random matrix and (right) symmetric random matrix of dimension  $n \times n$  with  $n = 1000$  where elements are drawn from the distribution  $\mathcal{N}(0, 1)/\sqrt{n}$ .

**Definition 2** (Stability). Let  $t_0 \geq 0$ , the equilibrium  $x^*$  is

- (i) *Stable*, if for all  $\epsilon > 0$  there exists a  $\delta(\epsilon, t_0) > 0$  such that  $\|x_0 - x^*\| \leq \delta$  implies  $\|\phi(t; t_0, x_0) - x^*\| \leq \epsilon$  for all  $t \geq t_0$ .
- (ii) *Attracting*, if there exists a  $\rho(t_0) > 0$  such that for all  $\eta > 0$  there exists a  $T(\eta, x_0, t_0)$  such that  $\|x_0 - x^*\| \leq \rho$  implies  $\|\phi(t; x_0, t_0) - x^*\| \leq \eta$  for all  $t \geq t_0 + T$ .
- (iii) *Uniformly Stable*, if the  $\delta$  in (i) is uniform in  $t_0$ , thus taking the form  $\delta(\epsilon)$ .
- (iv) *Uniformly Attracting*, if it is attracting where  $\rho$  does not depend on  $t_0$  and  $T(\eta, \rho)$  does not depend on  $x_0$  or  $t_0$ .
- (v) *Uniformly Asymptotically Stable* (UAS), if it is uniformly stable and uniformly attracting.
- (vi) *Uniformly Bounded*, if for all  $r > 0$  there exists a  $B(r)$  such that  $\|x_0 - x^*\| \leq r$  implies that  $\|s(t; t_0, x_0) - x^*\| \leq B$  for all  $t \geq t_0$ .
- (vii) *Uniformly Attracting in the Large*, if for all  $\rho > 0$  and  $\eta > 0$  there exists a  $T(\eta, \rho)$  such that  $\|x_0 - x^*\| \leq \rho$  implies  $\|s(t; x_0, t_0) - x^*\| \leq \eta$  for all  $t \geq t_0 + T$ .
- (viii) *Uniformly Asymptotically Stable in the Large* (UASL), if it is uniformly stable, uniformly bounded, and uniformly attracting in the large.
- (ix) *UAS in the Positive Orthant*, if it is uniformly stable, uniformly bounded, and uniformly attracting in the positive orthant.

The precise definition of UAS is important in the context of dynamical systems. If one is able to show that a given system is UAS, then it follows that the dynamics are stable in the presence of bounded disturbances as well. That is, if the dynamics in (T7) are UAS then for  $\|d(t)\| \leq \gamma$  where  $\gamma > 0$  is sufficiently small, the dynamics

$$\dot{x}(t) = f(x(t)) + d(t)$$

are uniformly stable. If the dynamics in (T7) are UASL then  $x(t)$  is bounded for all bounded  $d(t)$  and  $\gamma$  can be arbitrarily large. A detailed discussion regarding this fact can be found in [40, Definition 56.1 and Theorem 56.4] and a practical example in the context of adaptive systems can be found in [69].

For a linear dynamical system  $\dot{x}(t) = Ax(t)$  the uniform asymptotic stability (which is actually exponential) is verified if all of the eigenvalues of  $A$  have real parts less than zero. A well known theorem regarding the stability of linear systems, due to Lyapunov, is now stated.

**Theorem 4** (Lyapunov). *The eigenvalues of a real matrix  $A$  have all real parts less than zero if and only if there exists a  $P = P^T > 0$  such that  $A^T P + P A < 0$ .*

A stronger version of Theorem 4 will be needed when discussing the stability of Lotka-Volterra dynamics.

**Definition 3.** If there exists a diagonal positive matrix  $P$  such that  $A^T P + PA < 0$  then  $A$  is said to be *Diagonally Stable*.

## 4.2 Stability of Generalized Lotka Volterra Dynamics

Consider dynamics of the form

$$\dot{x}_i(t) = r_i x_i(t) + x_i(t) \sum_{j=1}^n a_{ij} x_j(t), \quad i = 1, \dots, n \quad (\text{T8})$$

where  $t \in [t_0, \infty)$  is time with  $t_0$  the initial time. The state vector is denoted  $x \in \mathbb{R}^n$  and defined as  $x = [x_1, x_2, \dots, x_n]^T$ . The linear terms are collected in the column vector  $r = [r_1, r_2, \dots, r_n]^T$  and  $A = (a_{ij})_{1 \leq i, j \leq n}$  captures the pair-wise interactions in the generalized Lotka-Volterra dynamics presented in (T8). The dynamics in (T8) can be compactly represented as

$$\dot{x}(t) = \text{diag}(x(t))(r + Ax(t)). \quad (\text{T9})$$

A discussion regarding the invertability of  $A$  is in order. This will become important in determining whether the system in (T8) has a unique non-trivial steady state. As discussed in the main text, the existence of the Verhulst terms  $a_{ii}x_i^2$  increases the likelihood that  $A$  is full rank. First consider the case where all  $a_{ij} = 0$  for all  $i \neq j$ , then a necessary and sufficient condition for  $A$  to be full rank is that all diagonal elements are non-zero.

If  $A$  is invertible then there exists a unique non-trivial steady state solution of the dynamics in (T8), denoted as  $x^* = -A^{-1}r$  [38]. We are interested in answering the following question. Given an ecological system of  $n$  species, is it possible to introduce another species and drive the system to any non-trivial steady state of our choosing? We will show that the answer is yes, and furthermore, if the original  $n$ -dimensional ecological system satisfies a diagonal stability condition, then we can design the interaction strengths for the introduced species so that the new  $n + 1$  dimensional ecological system is asymptotically stable for all initial conditions in the positive orthant. We then discuss equivalent results regarding steady state shift and stability when an arbitrary number of species is added.

Consider the following set of assumptions.

*Assumption 1.*  $A$  is invertible.

*Assumption 2.* For the dynamics in (T8) the steady state solution  $x^* \in \mathbb{R}_{>0}^n$ .

*Assumption 3.* The matrix  $A$  is diagonally stable, see Definition 3.

**Theorem 5** ([38, Theorem 1]). *If the system in (T8) satisfies Assumptions 1-3, then the steady state  $x^*$  is uniformly asymptotically stable for all initial conditions  $x_0 \in \mathbb{R}_{>0}^n$ .*

*Proof.* Let  $V(x, t) = 2 \sum_{i=1}^n p_i (x_i - x_i^* - x_i^* \log(x_i/x_i^*))$  be the Lyapunov candidate where  $p_i$  is the  $i$ -th diagonal element of a diagonal positive matrix  $P$  such that  $A^T P + PA < 0$ . Differentiating the Lyapunov candidate it follows that

$$\begin{aligned} \dot{V}(x, t) &= 2 \sum_{i=1}^n p_i \left( \dot{x}_i - x_i^* \frac{\dot{x}_i}{x_i} \right) \\ &= 2 \sum_{i=1}^n p_i (x_i - x_i^*) \frac{\dot{x}_i}{x_i} \\ &= 2 \sum_{i=1}^n p_i (x_i - x_i^*) \left( b_i + \sum_{j=1}^n a_{ij} x_j \right) \\ &= 2 \sum_{i=1}^n p_i (x_i - x_i^*) \sum_{j=1}^n a_{ij} (x_j - x_j^*) \\ &= (x - x^*)^T (A^T P + PA) (x - x^*). \end{aligned}$$

Thus the Lyapunov candidate is positive definite in  $x - x^*$  and its derivative is negative definite in  $x - x^*$ .  $\square$

*Remark 1.* Note that in the original work of Goh [38], stability in the positive orthant is denoted, but in fact what he proved is *uniform asymptotic* stability in the positive orthant. It is important to distinguish the two as stability only implies boundedness of trajectories and uniform asymptotic stability implies convergence to the equilibrium as well as robustness to bounded persistent disturbances.

### 4.3 Stability in the presence of new species

#### 4.3.1 Adding one new species

For the following examples we are interested in the  $m = n + 1$  dimensional dynamics

$$\dot{z}(t) = \text{diag}(z(t))(g + Fz(t)) \quad (\text{T10})$$

where

$$g = \begin{bmatrix} r \\ s \end{bmatrix}, \quad \text{and} \quad F = \begin{bmatrix} A & b \\ c^\top & d \end{bmatrix},$$

with  $A$  and  $r$  are as defined in (T8), and we have introduced the new elements  $s, d \in \mathbb{R}$ , and  $b, c \in \mathbb{R}^n$ . These dynamics represent the addition of one new species to the ecological system in (T8).

We now introduce an important property of diagonally stable matrices.

**Theorem 6** ([84, Theorem 3.1]). *Let  $A \in \mathbb{R}^{m \times m}$  be partitioned as*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

*where  $A_{11} \in \mathbb{R}^{(m-1) \times (m-1)}$ ,  $A_{12}, A_{21}^\top \in \mathbb{R}^{m-1}$ , and  $A_{22} < 0$ . Then  $A$  is diagonally stable if and only if  $A_{11}$  and the quantity  $(A_{11} - A_{12}A_{21}/A_{22})$  have a common diagonal Lyapunov function, i.e., there exists a positive diagonal  $P$  such that  $A_{11}^\top P + PA_{11} < 0$  and  $(A_{11} - A_{12}A_{21}/A_{22})^\top P + P(A_{11} - A_{12}A_{21}/A_{22}) < 0$ .*

**Lemma 1.** *For any steady state solution  $x^*$  of (T8) satisfying  $-Ax^* = r$  there exists  $b, c, d, s$  such that any  $z^* \in \mathbb{R}_{>0}^m$  can be made to be a steady state solution of (T10).*

*Proof.* Any steady solution of (T10) satisfies the relation  $g = -Fz^*$ , which when expanded denotes the following relation

$$\begin{bmatrix} r \\ s \end{bmatrix} = - \begin{bmatrix} A & b \\ c^\top & d \end{bmatrix} \begin{bmatrix} z_{1:n}^* \\ z_m^* \end{bmatrix}.$$

There are  $2n + 2$  degrees of freedom in the variables  $b, c, d, s$  and there are only  $n + 1$  constraints. The variable  $b$  is fixed by the top row of the above equation and can be expressed in closed form as

$$b = - \frac{r + Az_{1:n}^*}{z_m^*}. \quad (\text{T11})$$

Then for any  $c$  and  $d$ ,  $s$  can be chosen as

$$s = -c^\top z_{1:n}^* - dz_m^*. \quad \square \quad (\text{T12})$$

**Corollary 7.** *For any steady state solution  $x^*$  of (T8) satisfying  $-Ax^* = r$ , and given any  $c \in \mathbb{R}^n$  and  $d \in \mathbb{R}$  there exists  $b, s$  such that any  $z^* \in \mathbb{R}_{>0}^m$  can be made to be a steady state solution of (T10).*

**Theorem 8.** *For the dynamics in (T8) with  $p = 1$  satisfying Assumption 3 there exists  $b, c, d, s$  such that any  $z^* \in \mathbb{R}_{>0}^m$  can be made to be asymptotically stable for all initial conditions in the positive orthant.*

*Proof.* From Lemma 1 it follows that for any  $z^*$ ,  $c$  and  $d$  the elements  $s, b$  are fixed. Now we will show that for any  $c$  there exists a  $d$  such that  $z^*$  is asymptotically stable. We begin by assuming that  $b$  and  $s$  are fixed by (T11) and (T12). From Assumption 3 we know that  $A$  is diagonally stable, and thus there exists an  $\epsilon > 0$  and diagonal matrix  $P > 0$  such that

$A^\top P + PA \leq -\epsilon I$ , where  $I$  is an appropriately dimensioned identity matrix. Choosing  $d < 0$  and  $|d| > 2\lambda_{\max}(P)\|bc^\top\|/\epsilon$ , and noting that

$$(A - bc^\top/d)^\top P + P(A - bc^\top/d) \leq -\epsilon I + 2\lambda_{\max}(P)\|bc^\top\|/d$$

we can deduce that  $(A - bc^\top/d)^\top P + P(A - bc^\top/d) < 0$ . Thus  $(A - bc^\top/d)$  and  $A$  have a common diagonal Lyapunov function. This, in addition with the fact that  $d < 0$ , the conditions of Theorem 6 are satisfied and thus  $F$  is diagonally stable. Now, applying Theorem 5 to the dynamics in (T10) we deduce that  $z(t)$  can be made to be asymptotically stable for all  $z(t_0) \in \mathbb{R}_{>0}^m$ .  $\square$

### 4.3.2 Adding an arbitrary number of species

For the following examples we are interested in the  $m = n + p$  dimensional dynamics

$$\dot{z}(t) = \text{diag}(z(t))(g + Fz(t)) \quad (\text{T13})$$

where

$$g = \begin{bmatrix} r \\ s \end{bmatrix}, \quad \text{and} \quad F = \begin{bmatrix} A & B \\ C^\top & D \end{bmatrix},$$

with  $A$  and  $r$  are as defined in (T8), and we have introduced the new elements  $s \in \mathbb{R}^p$ ,  $D \in \mathbb{R}^{p \times p}$  and  $B, C \in \mathbb{R}^{n \times p}$ . These dynamics represent the addition of  $p$  new species to the ecological system in (T8).

**Theorem 9.** *For any steady state solution  $x^*$  of (T8) satisfying  $-Ax^* = r$  there exists  $B, C, D, s$  such that any  $z^* \in \mathbb{R}_{>0}^m$  can be made to be a steady state solution of (T13). Furthermore, if  $A$  is diagonally stable, then  $B, C, D, s$  can be chosen such that the system in (T13) is uniformly asymptotically stable in the positive orthant.*

*Proof.* Any steady solution of (T13) satisfies the relation  $g = -Fz^*$  which when expanded denotes the following relation

$$\begin{bmatrix} r \\ s \end{bmatrix} = - \begin{bmatrix} A & B \\ C^\top & D \end{bmatrix} \begin{bmatrix} z_{1:n}^* \\ z_{(n+1):m}^* \end{bmatrix}.$$

There are  $2np + p^2 + p$  degrees of freedom in the variables  $B, C, D, s$ . The variable  $B$  is fixed by the top row of the above equation and must satisfy the following relation

$$Bz_{(n+1):m}^* = -(r + Az_{1:n}^*).$$

There are  $n \times p$  degrees of freedom in the selection of  $B$  and only  $n$  constraints in the above equation. Thus such a  $B$  always exists for any  $r, A$ , and  $z^*$ . For any  $C$  and  $D$ ,  $s$  can be chosen as

$$s = -C^\top z_{1:n}^* - Dz_{(n+1):m}^*.$$

Finally, we show that with the extra degrees of freedom in  $C$  and  $D$  there always exists a diagonal  $P_1 > 0$  and  $P_2 > 0$  such

$$\begin{bmatrix} A & B \\ C^\top & D \end{bmatrix}^\top \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} + \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \begin{bmatrix} A & B \\ C^\top & D \end{bmatrix} < 0. \quad (\text{T14})$$

Given that by assumption there exists a diagonal  $P_1 > 0$  such  $\tilde{A} \triangleq A^\top P_1 + P_1 A$ , then by the Schur complement the inequality in (T14) holds if and only if

$$D^\top P_2 + 2P_2 D - (B^\top P_1 + P_2 C^\top) \tilde{A}^{-1} (CP_2 + P_1 B) < 0.$$

Given any  $A, B, C$  and positive diagonal  $P_1, P_2$  there always exists a  $D$  such that the above inequality holds.  $\square$

### 4.3.3 Removing an arbitrary number of species.

Diagonally stable matrices have a very special property that every principle minor is also diagonally stable, giving us the following definition and theorem.

**Definition 4.** Let  $L$  be a proper subset of  $N \triangleq \{1, 2, \dots, n\}$  then a *principle minor* of  $A \in \mathbb{R}^{n \times n}$  is the matrix obtained by omitting the columns and rows of  $A$  whose index appear in  $L$ .

**Theorem 10** ([19, Theorem 1]). *If  $A \in \mathbb{R}^{n \times n}$  is diagonally stable, then all principle minors of  $A$  are also diagonally stable.*

In the context of Lotka-Volterra dynamics this implies that if a given systems is diagonally stable, then even if an arbitrary numbers of species are removed from the system, the resulting system is still uniformly asymptotically stable in the positive orthant.

**Corollary 11.** *A necessary condition that  $A$  is diagonally stable is that all of its diagonal elements must be strictly negative.*

## 4.4 Robustness to disturbances.

We now consider three classes of Lotka-Volterra dynamics in the presence of disturbances, the first two are deterministic and the second is stochastic.

### 4.4.1 Deterministic Dynamics

In this section we will analyze the following two dynamical systems

$$\dot{x}(t) = \text{diag}(x(t))(r + Ax(t) + w(t)) \quad (\text{T15})$$

and

$$\dot{x}(t) = d(t) + \text{diag}(x(t))(r + Ax(t)) \quad (\text{T16})$$

where  $d(t)$  and  $w(t)$  are known to be a priori bounded. In (T15) the term  $w(t)$  represents uncertainty in the growth rates of the species. While in (T15) the term  $w$  is deterministic, in the since that it is bounded, in terms of the ecology literature this term is sometimes referred to as the stochastic effect. Note that we will treat this term in a purely stochastic since shortly, complete with Itô calculus. The term  $d(t)$  is a migration term. The following theorems show that in the presence of these disturbances, the systems are bounded for all initial conditions (in the positive orthant).

**Theorem 12.** *If the system in (T15) satisfies Assumptions 1-3 with  $\|d(t)\| \leq \alpha$  then the state  $x$  is uniformly bounded for all initial conditions  $x_0 \in \mathbb{R}_{>0}^n$ .*

*Proof.* The proof is given for two different scenarios. The first scenario (when the disturbance is small) uses the same Lyapunov function as was used in the analysis of the disturbance free dynamics. In the second scenario a new Lyapunov candidate is introduced. A few definitions are needed before the different scenarios are analyzed. Let  $P$  be a diagonal positive solution to the Lyapunov equation  $A^\top P + PA = -Q$ , where  $Q = Q^\top > 0$ , and  $p_i$  denotes the  $i$ -th diagonal element of  $P$ , just as in the disturbance free case. Let  $q_{\min}$  denote the minimum eigenvalue of  $Q$  and  $p_{\max}$  denote the maximum diagonal element in  $P$ . Scenario: (a) the compact set

$$\{x : \|x - x^*\| \leq 2p_{\max}\alpha/q_{\min}\}$$

does not intersect any of the  $n$ -axes; and (b) when the above compact set does intersect at least one of the  $n$ -axes.

We now address the stability of Scenario (a). Let

$$V(x) = 2 \sum_{i=1}^n p_i (x_i - x_i^* - x_i^* \log(x_i/x_i^*)) \quad (\text{T17})$$

be a Lyapunov candidate. Differentiating the Lyapunov candidate it follows that

$$\begin{aligned}
\dot{V}(x) &= 2 \sum_{i=1}^n p_i \left( \dot{x}_i - x_i^* \frac{\dot{x}_i}{x_i} \right) \\
&= 2 \sum_{i=1}^n p_i (x_i - x_i^*) \frac{\dot{x}_i}{x_i} \\
&= 2 \sum_{i=1}^n p_i (x_i - x_i^*) \left( b_i + d_i(t) + \sum_{j=1}^n a_{ij} x_j \right) \\
&= 2 \sum_{i=1}^n p_i (x_i - x_i^*) \left( d_i(t) + \sum_{j=1}^n a_{ij} (x_j - x_j^*) \right) \\
&= (x - x^*)^\top (A^\top P + PA)(x - x^*) + 2(x - x^*)^\top P d.
\end{aligned}$$

Recall that  $q_{\min}$  denotes the minimum eigenvalue of  $Q$  and  $p_{\max}$  denotes the maximum diagonal element in  $P$ , then it follows that

$$\dot{V} \leq -q_{\min} \|x - x^*\|^2 + 2p_{\max} \|x - x^*\| \alpha.$$

Thus for all  $\|x - x^*\| > 2p_{\max} \alpha / q_{\min}$ ,  $\dot{V} < 0$  and thus  $V(x(t)) < \infty$  for all  $t \geq t_0$ . It follows that for all  $x_0 \in \mathbb{R}_{>0}^n$  the state  $x(t)$  is bounded. Note that  $x(t)$  can never be negative due to the fact that the disturbance appears as  $x_i(t)d_i(t)$ . This completes the proof for Scenario (a).

Scenario (b) has to be treated differently due to the fact that if any of the  $x_i = 0$ , then it follows that for the Lyapunov candidate in (T17)  $V = \infty$ . This was not possible in Scenario (a), but is possible in Scenario (b). We define a new Lyapunov candidate

$$V(x) = 2 \sum_{i=1}^n v_i(x_i) \quad (\text{T18})$$

where

$$v_i(x) = \begin{cases} p_i(x_i - x_i^* - x_i^* \log(x_i/x_i^*)) & x_i^* \leq x_i \\ 0 & 0 \leq x_i < x_i^*. \end{cases}$$

For a general set of dynamics the above candidate function could not be a Lyapunov candidate. However, due to the special form of the Lotka-Volterra dynamics  $x(t) \in \mathbb{R}_{\geq 0}^n$  for all time regardless of the specific coefficients. Any population dynamic model should have this property, as negative abundances would make no sense. Differentiating  $V$  in (T18) we have that

$$\dot{V}(x) = 2 \sum_{i=1}^n \dot{v}_i(x_i)$$

where

$$\dot{v}_i(x) = \begin{cases} p_i(x_i - x_i^*) \left( d_i(t) + \sum_{j=1}^n a_{ij} (x_j - x_j^*) \right) & x_i^* \leq x_i \\ 0 & 0 \leq x_i < x_i^*, \end{cases}$$

which is continuous in  $x$ , note that  $p_i(x_i - x_i^*) \left( d_i(t) + \sum_{j=1}^n a_{ij} (x_j - x_j^*) \right) = 0$  when  $x_i = x_i^*$ .

We now define the index set for all species as

$$\mathcal{I} = \{i : 0 \leq i \leq n, i \in \mathbb{N}\} \quad (\text{T19})$$

and the set of indices for species abundances that are greater than  $x_i^*$  is defined as

$$\mathcal{S} = \{i : x_i > x_i^*\} \subset \mathcal{I}. \quad (\text{T20})$$

Using this notation we can write the derivative of the Lyapunov function as

$$\dot{V}(x) = 2 \sum_{i \in \mathcal{S}} p_i (x_i - x_i^*) \left( d_i + \sum_{j \in \mathcal{S}} a_{ij} (x_j - x_j^*) + \sum_{j \in \mathcal{S}^c} a_{ij} (x_j - x_j^*) \right).$$

The above equation can be rewritten in a more compact matrix-vector form as

$$\begin{aligned} \dot{V}(x) = & (x_S - x_S^*)^\top \left( A_{S,S}^\top P_{S,S} + P_{S,S} A_{S,S} \right) (x_S - x_S^*) + 2(x_S - x_S^*)^\top P_{S,S} d_S \\ & + 2(x_S - x_S^*)^\top P_{S,S} A_{S,S^c} (x_{S^c} - x_{S^c}^*). \end{aligned} \quad (\text{T21})$$

Using Theorem 10 we can exploit a very unique property of diagonal Lyapunov functions, and that is the diagonal stability of all principle submatrices. It is also easy to confirm that the same  $P$  matrix can be used for the sub-diagonal Lyapunov equations. Therefore, it follows that  $A_{S,S}^\top P_{S,S} + P_{S,S} A_{S,S} = -Q_{S,S} < 0$  for any  $S$ . Using the following definitions

$$\begin{aligned} q'_{\min}(\mathcal{S}) &= \min_i \lambda_i(Q_{S,S}) \\ p'_{\max}(\mathcal{S}) &= \max_i \lambda_i(P_{S,S}) \\ \beta'(\mathcal{S}) &= \|A_{S,S^c}\|, \end{aligned} \quad (\text{T22})$$

the equality in (T21) can be bounded as

$$\dot{V}(x) \leq -q'_{\min}(\mathcal{S}) \|x_S - x_S^*\|^2 + 2p'_{\max}(\mathcal{S}) \|x_S - x_S^*\| (\alpha + \beta'(\mathcal{S}) \|x_{S^c} - x_{S^c}^*\|).$$

Given our definition of  $\mathcal{S}$  it follows that  $\|x_{S^c} - x_{S^c}^*\| \leq \|x_{S^c}^*\| \leq \|x^*\|$ . This allows the above inequality to be further simplified as

$$\dot{V}(x) \leq -q'_{\min}(\mathcal{S}) \|x_S - x_S^*\|^2 + 2p'_{\max}(\mathcal{S}) \|x_S - x_S^*\| (\alpha + \beta'(\mathcal{S}) \|x^*\|). \quad (\text{T23})$$

The bound in (T23) is still not sufficient to address stability as the bounds depend on the set of indices in  $\mathcal{S}$ . In a two step process we obtain uniform constants and then introduce a distance function that is independent of  $\mathcal{S}$ . We will now give uniform analogs to the bounds  $q'_{\min}(\mathcal{S})$ ,  $p'_{\max}(\mathcal{S})$ ,  $\beta'(\mathcal{S})$  by taking the appropriate supremum or infimum over the powerset  $2^{\mathcal{I}}$  of all subsets  $\mathcal{S} \subset \mathcal{I}$ . The powerset of interest is finite and thus the following are well defined

$$\begin{aligned} \bar{q} &= \inf_{\mathcal{S} \subset \mathcal{I} \setminus \emptyset} q'_{\min}(\mathcal{S}) \\ \bar{p} &= \sup_{\mathcal{S} \subset \mathcal{I}} p'_{\max}(\mathcal{S}) \\ \bar{\beta} &= \sup_{\mathcal{S} \subset \mathcal{I}} \beta'(\mathcal{S}). \end{aligned} \quad (\text{T24})$$

Using the uniform bounds above the inequality in (T23) can be replaced by

$$\dot{V}(x) \leq -\bar{q} \|x_S - x_S^*\|^2 + 2\bar{p} \|x_S - x_S^*\| (\alpha + \bar{\beta} \|x^*\|). \quad (\text{T25})$$

We now define a distance metric between an element  $y \in \mathbb{R}^n$  and a compact set  $\mathcal{A} \in \mathbb{R}^n$  as

$$d(y, \mathcal{A}) \triangleq \inf\{\|y - z\| : z \in \mathcal{A}\}. \quad (\text{T26})$$

Given that  $\mathcal{A}$  is compact, such a  $z \in \mathcal{A}$  always exists. Our compact set of interest is defined as

$$\mathcal{X} = \{x : 0 \leq x_i \leq x_i^*\}. \quad (\text{T27})$$

Using the compact set defined just above, the inequality in (T25) can be equivalently rewritten as

$$\dot{V}(x) \leq -\bar{q} d(x, \mathcal{X})^2 + 2\bar{p} d(x, \mathcal{X}) (\alpha + \bar{\beta} \|x^*\|).$$

Thus for all  $d(x, \mathcal{X}) > 2\bar{p} (\alpha + \bar{\beta} \|x^*\|) / \bar{q}$ ,  $\dot{V} < 0$ . Therefore  $V(x(t))$  is bounded for all  $t \geq t_0$ .  $\square$

**Theorem 13.** *If the system in (T16) satisfies Assumptions 1-3 with  $\|d(t)\| \leq \alpha$  and furthermore we exclude the possibility for the disturbance to generate negative state values<sup>2</sup>, then the state  $x$  is uniformly bounded for all initial conditions  $x_0 \in \mathbb{R}_{\geq 0}^n$ .*

<sup>2</sup>This is not a restrictive assumption in the context of ecology as a disturbance can never result in the creation of a negative population.

*Proof.* Consider the Lyapunov candidate in (T18) and differentiating along the system dynamics in (T16) it follows that

$$\dot{V}(x) = 2 \sum_{i \in \mathcal{S}} p_i (x_i - x_i^*) \left( d_i/x_i + \sum_{j \in \mathcal{S}} a_{ij} (x_j - x_j^*) + \sum_{j \in \mathcal{S}^c} a_{ij} (x_j - x_j^*) \right).$$

where  $\mathcal{S} \subset I$  was defined in (T19) and (T20). Rearranging terms slightly with regard to  $d_i/x_i$  it follows that

$$\dot{V}(x) = 2 \sum_{i \in \mathcal{S}} p_i (x_i - x_i^*) \left( \sum_{j \in \mathcal{S}} a_{ij} (x_j - x_j^*) + \sum_{j \in \mathcal{S}^c} a_{ij} (x_j - x_j^*) \right) + 2 \sum_{i \in \mathcal{S}} d_i p_i (1 - x_i^*/x_i).$$

Given our definition of  $\mathcal{S}$  it follows that  $(1 - x_i^*/x_i) \leq 1$  when  $i \in \mathcal{S}$ . Therefore it follows that

$$\dot{V}(x) \leq 2 \sum_{i \in \mathcal{S}} p_i (x_i - x_i^*) \left( \sum_{j \in \mathcal{S}} a_{ij} (x_j - x_j^*) + \sum_{j \in \mathcal{S}^c} a_{ij} (x_j - x_j^*) \right) + 2 \sum_{i \in \mathcal{S}} |d_i| p_i.$$

The above inequality can be rewritten in a more compact matrix-vector form as

$$\begin{aligned} \dot{V}(x) &\leq (x_{\mathcal{S}} - x_{\mathcal{S}}^*)^\top \left( A_{\mathcal{S},\mathcal{S}}^\top P_{\mathcal{S},\mathcal{S}} + P_{\mathcal{S},\mathcal{S}} A_{\mathcal{S},\mathcal{S}} \right) (x_{\mathcal{S}} - x_{\mathcal{S}}^*) + 2 \|P_{\mathcal{S},\mathcal{S}}\| \|d_{\mathcal{S}}\| \\ &\quad + 2 (x_{\mathcal{S}} - x_{\mathcal{S}}^*)^\top P_{\mathcal{S},\mathcal{S}} A_{\mathcal{S},\mathcal{S}^c} (x_{\mathcal{S}^c} - x_{\mathcal{S}^c}^*). \end{aligned}$$

Using the bounds in (T22) the above inequality can be reduced to

$$\dot{V}(x) \leq -q'_{\min}(\mathcal{S}) \|x_{\mathcal{S}} - x_{\mathcal{S}}^*\|^2 + 2p'_{\max}(\mathcal{S}) \|x_{\mathcal{S}} - x_{\mathcal{S}}^*\| \beta'(\mathcal{S}) \|x_{\mathcal{S}^c} - x_{\mathcal{S}^c}^*\| + 2p'_{\max}(\mathcal{S}) \alpha.$$

Given our definition of  $\mathcal{S}$  it follows that  $\|x_{\mathcal{S}^c} - x_{\mathcal{S}^c}^*\| \leq \|x_{\mathcal{S}^c}^*\| \leq \|x^*\|$ . This allows the above inequality to be further simplified as

$$\dot{V}(x) \leq -q'_{\min}(\mathcal{S}) \|x_{\mathcal{S}} - x_{\mathcal{S}}^*\|^2 + 2p'_{\max}(\mathcal{S}) \|x_{\mathcal{S}} - x_{\mathcal{S}}^*\| \beta'(\mathcal{S}) \|x^*\| + 2p'_{\max}(\mathcal{S}) \alpha. \quad (\text{T28})$$

Using the definitions in (T24) and the set distance function in (T26) with the compact set  $\mathcal{X}$  defined in (T27) the above inequality can be written as

$$\dot{V}(x) \leq -\bar{q} d(x, \mathcal{X})^2 + 2\bar{p} d(x, \mathcal{X}) \bar{\beta} \|x^*\| + 2\bar{p} \alpha$$

Rewriting the above inequality as

$$\dot{V}(x) \leq -\frac{\bar{q}}{2} d(x, \mathcal{X})^2 - \frac{\bar{q}}{2} \left( d(x, \mathcal{X})^2 - 4 \frac{\bar{p}}{\bar{q}} d(x, \mathcal{X}) \bar{\beta} \|x^*\| \right) + 2\bar{p} \alpha$$

and completing the square with respect to the middle term  $\frac{\bar{q}}{2} \left( d(x, \mathcal{X})^2 - 4 \frac{\bar{p}}{\bar{q}} d(x, \mathcal{X}) \bar{\beta} \|x^*\| \right)$  the following inequality holds

$$\dot{V}(x) \leq -\frac{\bar{q}}{2} d(x, \mathcal{X})^2 - \frac{\bar{q}}{2} \left( d(x, \mathcal{X}) - 2 \frac{\bar{p}}{\bar{q}} \bar{\beta} \|x^*\| \right)^2 + 2\bar{p} \alpha + 2 \frac{\bar{p}^2 \bar{\beta}^2 \|x^*\|^2}{\bar{q}}$$

Noting that the term  $-\frac{\bar{q}}{2} \left( d(x, \mathcal{X}) - 2 \frac{\bar{p}}{\bar{q}} \bar{\beta} \|x^*\| \right)^2 \leq 0$  it follows that

$$\dot{V}(x) \leq -\frac{\bar{q}}{2} d(x, \mathcal{X})^2 + 2\bar{p} \alpha + 2 \frac{\bar{p}^2 \bar{\beta}^2 \|x^*\|^2}{\bar{q}} \quad (\text{T29})$$

From the inequality in (T29) it follows that

$$d(x, \mathcal{X}) > 2 \sqrt{\frac{\bar{p} \alpha}{\bar{q}} + \frac{\bar{p}^2 \bar{\beta}^2 \|x^*\|^2}{\bar{q}^2}}$$

implies  $\dot{V} < 0$ . Therefore,  $x(t)$  is uniformly bounded and asymptotically converges to the compact set

$$d(x, \mathcal{X}) \leq 2 \sqrt{\frac{\bar{p} \alpha}{\bar{q}} + \frac{\bar{p}^2 \bar{\beta}^2 \|x^*\|^2}{\bar{q}^2}}. \quad \square$$



#### 4.4.2 Stochastic Dynamics

In this section we wish to analyze a stochastic differential equation that is similar to (T15), but we no longer make the assumption that the disturbance is bounded, instead we assume that the disturbance appears as a Brownian motion  $w(t) \in \mathbb{R}^n$ , resulting in the following differential equation,

$$dx = \text{diag}(x)(r + Ax) dt + c \text{diag}(x) dw. \quad (\text{T30})$$

The variable  $c \in \mathbb{R}$  will be a constant used to scale the square root of the variance of the brownian motion. It will be shown that  $x(t)$  converges in a probabilistic sense to a compact set which is proportional to  $c$ . The analysis of stochastic differentials is significantly more challenging than their deterministic counterparts. Before stating the stochastic version of the stability result, we first need to formally define filtration, Brownian motion, and give a key result of Itô. In this work the linear ordered set of interest will always be time  $t \in [t_0, \infty)$ . All of the following definitions are given in a less general context compared to their original presentation [24].

**Definition 5** (Stochastic Process, [24, Part 2 §I.8]). Let  $(\Omega, \mathcal{F}, \mathbf{P})$  be a probability space. A *stochastic process* on  $(\Omega, \mathcal{F}, \mathbf{P}, [t_0, \infty))$  is a family of random variables  $\{y(t), t \in [t_0, \infty)\}$  with map  $(t, \omega) \mapsto y(t, \omega)$  from  $[t_0, \infty) \times \Omega$  to  $R$ , the variable space of interest. Given that each instance in time  $y(t)$  is a random variable, the variable space  $R = (R, \mathcal{R})$  is necessarily a measurable space.

**Definition 6** (Filtration, [24, Part 2 §I.1]). Let  $(\Omega, \mathcal{F})$  be a measurable space and  $t \in [t_0, \infty)$  is time, then a *filtration* of that measurable space is a map  $t \mapsto \mathcal{F}(t)$  of increasing sub  $\sigma$  algebras such that  $\mathcal{F}(s) \subset \mathcal{F}(t) \subset \mathcal{F}$  when  $s \leq t$ .

**Definition 7** (Adaptation, [24, Part 2 §I.1]). Let  $\{y(t), t \in [t_0, \infty)\}$  be a stochastic process from a *filtered measure space*  $\{\Omega, \mathcal{F}, \mathcal{F}(\cdot)\}$  into the measurable space  $\{R, \mathcal{R}\}$ . The process is *adapted* to  $\mathcal{F}(\cdot)$  if for each  $t$ ,  $y(t) : (\Omega, \mathcal{F}(t)) \rightarrow (R, \mathcal{R})$  is measurable.

**Definition 8** (Progressively Measurable, [24, Part 2 §I.2]). Let  $\{y(t), \mathcal{F}(t), t \in [t_0, \infty)\}$  be an adapted stochastic process. The function  $y : [t_0, \infty) \times \Omega \rightarrow R$ , where  $(R, \mathcal{R})$  is a measurable space, is deemed *progressively measurable* if

$$y : ([t_0, t] \times \Omega, \text{Borel}([t_0, t]) \times \mathcal{F}(t)) \rightarrow (R, \mathcal{R})$$

is measurable for any  $t \geq t_0$ .

Let us pause now and discuss these definitions. Defining a stochastic processes is rather obvious. We extended the definition of a random variable to incorporate time. At each fixed instance a stochastic process is nothing but a random variable. The extra definitions and the progression from filtrations, adaptations, and progressive measurability, are needed so that we can better understand what is needed to perform the following integration [88, Remark 7.1.1]. Let  $\{y(\cdot), \mathcal{F}(\cdot)\}$  be an adapted stochastic process defined as before where  $y : [t_0, \infty) \times \Omega \rightarrow R$ . Let  $f : R \rightarrow \mathbb{R}$  be bounded and  $\mathcal{R}$  measurable, then the map  $(t, \omega) \mapsto z = \int_{t_0}^t f(y(\tau, \omega)) d\tau$  need not be adapted. However if in addition to being adapted the stochastic process is progressively measurable, then  $z$  will be progressively measurable as well. This process is a key necessary ingredient when studying stochastic differential equations as it can be inherited. Next we formally define a Brownian motion and give a sufficient condition for the adaptation so that the progressive measurability condition is satisfied.

**Definition 9** (Brownian, [24, Part 2 §VII.2]). A *Brownian motion* is an adapted stochastic process  $\{w(\cdot), \mathcal{F}(\cdot)\}$  from the filtered probability space  $(\Omega, \mathcal{F}, \mathbf{P}, \mathcal{F}(t), t \in [t_0, \infty))$  into the measurable space  $(\mathbb{R}^n, \text{Borel}(\mathbb{R}^n))$

- that is *Markovian*, i.e. when  $s < t$  and  $\mathcal{A} \in \text{Borel}(\mathbb{R}^n)$  it follows that

$$\mathbf{P}(w(t) \in \mathcal{A} | \mathcal{F}(s)) = \mathbf{P}(w(t) \in \mathcal{A} | w(s))$$

almost surely.

- with a stationary stochastic transition function (the transition function is independent of the current time of the process) with density  $\rho(t, \psi - \xi)$  defined on  $\mathbb{R} \times \mathbb{R}^n$  relative to the  $n$ -dimensional Lebesgue measure, where  $\rho$  is defined as

$$\rho(t, \psi; \sigma, t_0) = \begin{cases} (2\pi\sigma^2 t)^{-n/2} \exp \frac{-|\psi|^2}{2\sigma^2 t} & \text{if } t > t_0 \\ 0 & \text{if } t \leq t_0, \end{cases}$$

where  $t \in \mathbb{R}$  is time and  $\psi \in \mathbb{R}^n$  is the space variable.

- that is almost surely continuous

We are almost ready to address the problem at hand, however we still need to show how we can ensure that the Brownian motion is progressively measurable. Unfortunately we need a few more definitions and then we will show that Brownian motions can be naturally adapted so as to have the progressive measurability property.

**Definition 10** (Right Continuous Filtration, [24, Part 2 §I.1]). Let  $(\Omega, \mathcal{F}, \mathcal{F}(t), t \in [t_0, \infty))$  be a filtered measurable space, and define  $\mathcal{F}^+(t) = \bigcap_{s>t} \mathcal{F}(s)$  for all  $t \in [t_0, \infty)$ . The filtration is *right-continuous* if  $\mathcal{F}(\cdot) = \mathcal{F}^+(\cdot)$ .

**Theorem 14.** Let  $\{y(\cdot), \mathcal{F}(\cdot)\}$  be a Brownian motion into  $\mathbb{R}^n$  with respect to time  $t \in [t_0, \infty)$ , and if  $\mathcal{F}(t)$  is generated by the null sets and  $\sigma(y(s); s \leq t)$ , the smallest sigma algebra for which all  $y(s)$  are measurable for all  $s \in [t_0, t]$ , then  $\mathcal{F}(\cdot) = \mathcal{F}^+(\cdot)$

*Proof.* see [24, Part 2, §VI, Theorem 8] □

This theorem implies that one can assume without loss of generality that the filtration is right-continuous. That is, there is a natural way to construct them for any brownian motion.

**Theorem 15.** Let  $\{y(\cdot), \mathcal{F}(\cdot)\}$  be a Brownian motion with right-continuous filtration containing the null sets, then the Brownian motion is progressively measurable.

*Proof.* see [66, Part A Theorem 47] □

As stated before, this progressive measurability is necessary when discussing the solutions to stochastic differential equations as it allows for integrals containing stochastic processes to inherit the progressively measurable property. We now state a classic result do to Itô [42]. Our version follows from [24, Part 2 §VIII.12].

**Lemma 2.** Let  $\{w(\cdot), \mathcal{F}(\cdot)\}$  be a Brownian motion into  $\mathbb{R}^n$  with a right-continuous filtration containing the null sets. Consider the dynamics  $x(t) \in \mathbb{R}^n$  given by

$$dx = \mu dt + \sigma dw$$

where  $t \in [t_0, \infty)$ ,  $\mu(x) \in \mathbb{R}^n$  is locally Lipschitz in  $x$ , and  $\sigma(x) \in \mathbb{R}^{n \times n}$  is globally Lipschitz in  $x$ . Assuming  $(x, t) \mapsto f(x, t) \in \mathbb{R}$  is twice differentiable and continuous in terms of  $x$  (class  $C^2$  with respect to  $x$ ) and once differentiable and continuous in terms of  $t$  (class  $C^1$  with respect to  $t$ ), then

$$df = \frac{\partial f}{\partial t} dt + (\nabla_x f)^\top \mu dt + \frac{1}{2} \text{tr}(\sigma \text{Hess}_x(f) \sigma) dt + (\nabla_x f)^\top \sigma dw.$$

*Remark 2.* In most constructions it is assumed that  $\mu$  is globally Lipschitz. Indeed this assumption was made in one of Itô's original papers [43]. A detailed discussion regarding the existence of solutions when only locally Lipschitz conditions are assumed can be found in [78]. Exploiting the stability of our dynamics the existence and uniqueness of solutions can be deduced directly from the original work of Itô however. In our analysis we will show that for the dynamics of interest the state is uniformly bounded almost surely, a formal definition is given just below this remark. Therefore, we can define equivalent dynamics

$$d\bar{x} = \bar{\mu} dt + \sigma dw$$

where  $\bar{\mu} = \mu$  on the compact set of interest, and zero outside this compact set. Given that  $\mu$  is locally Lipschitz it follows that the function  $\bar{\mu}$  with compact support is by definition globally Lipschitz. Solutions  $\bar{x}$  exist, are unique, and are uniformly bounded almost surely. This then implies the existence and uniqueness of  $x$  for the dynamics  $dx = \mu dt + \sigma dw$  on the same compact set which  $x(\cdot)$  remains in, almost surely.

**Definition 11.** The state  $x(t; t_0, x_0)$  as a solution to the difference equation in (T30) is *uniformly bounded with probability one* if for ever  $r > 0$  there exists a  $B(r) > 0$  such that  $\|x(t_0)\| \leq r$  implies

$$\mathbf{P} \left( \sup_t \|x(t)\| \leq B(r) \right) = 1$$

for all  $t \geq t_0$ . An equivalent statements would be  $x(t)$  is *uniformly bounded almost surely*.

**Theorem 16.** Consider the dynamics in (T30) for  $x(t) \in \mathbb{R}^n$ . Let assumptions 1-3 hold with  $\{w(\cdot), \mathcal{F}(\cdot)\}$  a Brownian motion into  $\mathbb{R}^n$  with a right-continuous filtration containing the null set. The following then hold

1. The state variable  $x(\cdot)$  is uniformly bounded in expectation,
2.  $x(\cdot)$  is uniformly bounded with probability one, and
3.  $x(\cdot)$  asymptotically converges to the compact set

$$\mathcal{D} \triangleq \left\{ x : \|x - x^*\| \leq c \sqrt{\frac{nx_{\max}^* p_{\max}}{q_{\min}}} \right\} \quad (\text{T31})$$

with probability one. Stated more precisely,

$$\mathbf{P} \left( \lim_{t \rightarrow \infty} x(t) \in \mathcal{D} \right) = 1. \quad (\text{T32})$$

*Sketch of the proof.* This proof only outlines the analysis when the compact set  $\mathcal{D}$  does not intersect any of the  $n$ -axes. The more complicated scenario when this is not the case can be handled just as it was in the proof of Theorem 12. Consider the Lyapunov candidate  $V(x) = 2 \sum_{i=1}^n p_i (x_i - x_i^* - x_i^* \log(x_i/x_i^*))$ . Taking the differential along the lines of Lemma 2 (Itô) results in

$$\begin{aligned} dV = (\nabla_x V)^\top \text{diag}(x)(r + Ax) dt + \frac{c^2}{2} \text{tr}(\text{diag}(x) \text{Hess}_x(V) \text{diag}(x)) dt \\ + (\nabla_x V)^\top c \text{diag}(x) dw \end{aligned} \quad (\text{T33})$$

Recall from the steps in the proof of Theorem 12 that

$$(\nabla_x V)^\top \text{diag}(x)(r + Ax) = -(x - x^*)^\top Q(x - x^*)$$

where  $A^\top P + PA = -Q$ . Thus giving

$$\begin{aligned} dV = -(x - x^*)^\top Q(x - x^*) dt + \frac{c^2}{2} \text{tr}(\text{diag}(x) \text{Hess}_x(V) \text{diag}(x)) dt \\ + (\nabla_x V)^\top c \text{diag}(x) dw \end{aligned} \quad (\text{T34})$$

Next, note that  $[\text{Hess}_x(V)]_{ii} = 2 \frac{p_i x_i^*}{x_i^2}$  and  $[\text{Hess}_x(V)]_{ij} = 0$  when  $i \neq j$ . Substitution into the above equation results in

$$dV = -(x - x^*)^\top Q(x - x^*) dt + c^2 \sum_i p_i dt + (\nabla_x V)^\top c \text{diag}(x) dw \quad (\text{T35})$$

Note that  $x(t)$  and  $dw(t)$  are independent,  $x(t)$  is only dependent on  $dw(s)$  when  $s < t$ . Therefore  $\mathbf{E}((\nabla_x V)^\top c \text{diag}(x) dw) = 0$ , given that for a Brownian motion  $\mathbf{E}dw = 0$ .<sup>3</sup> Therefore

$$\frac{\mathbf{E} dV}{dt} \leq -q_{\min} \|x - x^*\|^2 + c^2 n x_{\max}^* p_{\max} \quad (\text{T36})$$

where  $q_{\min}$  is the minimum eigenvalue of  $Q$ ,  $p_{\max}$  is the maximum diagonal element in  $P$ , and  $x_{\max}^*$  is the value in the vector  $x^*$ . Therefore,

$$\|x - x^*\| > c \sqrt{\frac{nx_{\max}^* p_{\max}}{q_{\min}}} \implies \frac{\mathbf{E} dV}{dt} < 0. \quad (\text{T37})$$

<sup>3</sup>In order to address this rigorously we would rewrite the expression in (T35) in integral form. Then we would integrate only up to a stopping time which we designed so that  $(\nabla_x V)^\top c \text{diag}(x) < \infty$  up to the time of interest. Then we would use the fact that  $\mathbf{E}gw = 0$  for any bounded  $g$ . The bound used to generate the stopping time would then be relaxed and the desired result would be obtained as  $t \rightarrow \infty$  using Fatou's Lemma and monotone convergence. A similar procedure is carried out in [21, (4.3)-(4.5)]

Given that  $\mathbf{E}dV/dt \leq 0$  outside a compact set it follows that  $\mathbf{E}V$  is uniformly bounded in terms of the initial condition  $x(t_0)$  [50, Lemma 5.4]. Given that  $V(x) \geq 0$  is continuous in  $x$  and convex it follows from Jensen's inequality that  $\mathbf{E}x$  is uniformly bounded as well [47]. Jensen's inequality will be used without reference from this point forward. Claim 1 has been proven.

We now move on to the second claim, namely that the dynamics are uniformly bounded with probability one. We address this claim with sub scenarios: (a)  $x(t_0) \notin \mathcal{D}$ , and (b)  $x(t_0) \in \mathcal{D}$ . Under scenario (a) we will now show that  $\mathbf{P}(\sup V(x(t)) - V(x(t_0)) \leq \delta_1) = 1$  for any  $\delta_1 \in (0, \infty)$ . The previous statement is equivalent to  $\mathbf{P}(\sup V(x(t)) - V(x(t_0)) > \delta_1) = 0$ . This statement will be proved by contradiction. Assume that there exists a  $\delta_2 \in (0, 1]$  such that  $\mathbf{P}(\sup V(x(t)) - V(x(t_0)) > \delta_1) \geq \delta_2$ . Using Markov's inequality it follows that

$$\mathbf{E} \sup V(x(t)) - V(x(t_0)) \geq \delta_1 \delta_2.$$

This statement however contradicts (T37) which states that outside the compact set of interest the expected value of the Lyapunov candidate is strictly decreasing. Therefore it follows that  $\mathbf{P}(\sup V(x(t)) - V(x(t_0)) > \delta_1) = 0$  which is equivalent to the claim that  $\mathbf{P}(\sup V(x(t)) - V(x(t_0)) \leq \delta_1) = 1$ . Application of Jensen's inequality then implies that  $x$  is uniformly bounded with probability one for sub-scenario (a).

We now address sub-scenario (b) where it is assumed that  $x(t_0) \in \mathcal{D}$ . If  $x(\cdot)$  remains in  $\mathcal{D}$  for all time, then we are done and the state is uniformly bounded. Therefore, assume at some time  $t_1$  the state  $x(t_1) \notin \mathcal{D}$ . This sub-scenario is now equivalent to sub-scenario (a) with time shifted. Therefore, it follows that  $x$  is uniformly bounded with probability 1 in sub-scenario (b) as well. This completes the proof of claim 2.

We approach claim 3 following a method proposed in the proof of [21, Theorem 2.1]. Consider the probability of three mutually exclusive scenarios, just as in [21, Theorem 2.1]

$$\begin{aligned} p_1 &= \mathbf{P} \left( \limsup_{t \rightarrow \infty} d(x(t), \mathcal{D}) = 0 \right) \\ p_2 &= \mathbf{P} \left( \liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) > 0 \right) \\ p_3 &= \mathbf{P} \left( \liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) = 0 \text{ and } \limsup_{t \rightarrow \infty} d(x(t), \mathcal{D}) > 0 \right) \end{aligned}$$

We wish to prove that  $p_1 = 1$  and  $p_2, p_3 = 0$ .

We first prove that  $p_2 = 0$ . This will be proved by contradiction. Assume  $p_2 = \epsilon_1$  where  $\epsilon_1 \in (0, 1]$ , then there exists an  $\epsilon_2$  such that  $\mathbf{P}(\liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) \geq \epsilon_2) = \epsilon_1$ . Using Markov's inequality it follows that

$$\frac{1}{\epsilon_2} \mathbf{E} \liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) \geq \epsilon_1.$$

Multiplying both sides by  $\epsilon_2$  it follows that

$$\mathbf{E} \liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) \geq \epsilon_1 \epsilon_2.$$

From the definition of  $\liminf$  it follows that for any  $\epsilon_3 \in (0, \epsilon_1 \epsilon_2)$  there exist a finite  $T_1 \in [t_0, \infty)$  such that

$$\mathbf{E} d(x(t), \mathcal{D}) \geq \epsilon_3 \text{ for all } t \geq T_1. \quad (\text{T38})$$

Therefore, for  $t \geq T_1$  it follows that  $\mathbf{E}dV/dt < 0$ . Given that  $V(x^*) = 0$  and  $V(x)$  is strictly increasing away from  $x = x^*$  and tends to infinity for large  $x$  in the positive orthant it follows that  $\mathbf{E}\|x - x^*\|$  is strictly decreasing in time. Given that  $\mathcal{D}$  is centered at  $x^*$ , there exists  $T_2$  such that

$$\mathbf{E} d(x(t), \mathcal{D}) < \epsilon_3 \text{ for all } t \geq T_2.$$

which contradicts (T38). Therefore  $p_2 = 0$ . That is

$$\mathbf{P} \left( \liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) > 0 \right) = 0.$$

We now establish that  $p_3 = 0$ . As before we achieve this by contradiction. For any  $\epsilon_5 > 0$  assume

$$\mathbf{P} \left( \liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) = 0 \text{ and } \limsup_{t \rightarrow \infty} d(x(t), \mathcal{D}) \geq \epsilon_5 \right) \neq 0. \quad (\text{T39})$$

For any  $\epsilon_6$  and  $\epsilon_7$  such that  $0 < \epsilon_7 < \epsilon_6 < \epsilon_5$ , consider the *stopping times*<sup>4</sup>  $T'_i : \Omega \rightarrow [t_0, \infty)$ ,  $T''_i : \Omega \rightarrow [t_0, \infty)$  with  $i \in \mathbb{N}_{>0}$ , where

$$\begin{aligned} T'_i &= \inf \{t \geq T''_{i-1} : d(x(t), \mathcal{D}) \leq \epsilon_7\} \\ T''_i &= \inf \{t \geq T'_i : d(x(t), \mathcal{D}) \geq \epsilon_6\}. \end{aligned}$$

From (T39) and the stopping times defined above, it follows that

$$\lim_{i \rightarrow \infty} T'_i = \infty \text{ and } \lim_{i \rightarrow \infty} T''_i = \infty$$

In order to continue with this proof by contradiction we need to obtain a lower bound on the expectation  $T'_{i+1} - T''_i$ . Indeed if this can be done, then we expect the solutions to the differential equation to spend a non negligible amount of time within a domain of the state space where the Lyapunov function will be decreasing. Thus we can see how this may lead to a contradiction, for how can trajectories be expected to spend an infinite amount of time in a domain where the Lyapunov function is always decreasing. Following the same procedures as in the proof of the bound in [21, (2.27)] it can be shown that there exists an  $\epsilon_8 > 0$  such that

$$\mathbf{E}(T'_{i+1} - T''_i \mid \mathcal{F}(T''_i)) \geq \epsilon_8.$$

Using the bound above and the definition of  $\mathcal{D}$  in (T31) it follows that

$$\begin{aligned} \mathbf{E} \left( \int_{T''_i}^{T'_{i+1}} q_{\min} \|x(\tau) - x^*\|^2 - c^2 n x_{\max}^* p_{\max} d\tau \mid \mathcal{F}(T''_i) \right) \\ \geq \left( q_{\min} c \sqrt{\frac{n x_{\max}^* p_{\max}}{q_{\min}}} \epsilon_6 + q_{\min} \epsilon_6^2 \right) \epsilon_8 \end{aligned}$$

From the bound in (T36) it follows that

$$\begin{aligned} V(x(t_0)) &\geq \mathbf{E} \int_{t_0}^{\infty} q_{\min} \|x(\tau) - x^*\|^2 - c^2 n x_{\max}^* p_{\max} d\tau \\ &\geq \sum_{i=1}^{\infty} \left( q_{\min} c \sqrt{\frac{n x_{\max}^* p_{\max}}{q_{\min}}} \epsilon_6 + q_{\min} \epsilon_6^2 \right) \epsilon_8 \mathbf{P}(T''_i < \infty). \end{aligned}$$

Noting that  $V(x(t_0)) < \infty$ , it then follows from the Borel-Cantelli Lemma [23, Chapter III, Theorem 1.2] that  $\mathbf{P}(\exists N < \infty \text{ s.t. } \forall i \geq N, T''_i < \infty) = 0$  [21, (2.28)-(2.31)]. This implies that  $\mathbf{P}(\limsup_{t \rightarrow \infty} d(x(t), \mathcal{D}) \geq \epsilon_5) = 0$  which contradicts (T39). Therefore  $p_3 = 0$ , that is

$$\mathbf{P} \left( \liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) = 0 \text{ and } \limsup_{t \rightarrow \infty} d(x(t), \mathcal{D}(0)) > 0 \right) = 0.$$

Finally, given that the events associated with  $p_1$ ,  $p_2$ , and  $p_3$  are nonintersecting, it follows that  $p_1 = 1$ . Summarizing, we have shown that

$$\begin{aligned} p_1 &= \mathbf{P} \left( \limsup_{t \rightarrow \infty} d(x(t), \mathcal{D}) = 0 \right) = 1 \\ p_2 &= \mathbf{P} \left( \liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) > 0 \right) = 0 \\ p_3 &= \mathbf{P} \left( \liminf_{t \rightarrow \infty} d(x(t), \mathcal{D}) = 0 \text{ and } \limsup_{t \rightarrow \infty} d(x(t), \mathcal{D}(0)) > 0 \right) = 0, \end{aligned}$$

and thus claim 2 of the theorem has been proven. We wish to reiterate the fact that this sketch barrows heavily from [21]. In most instances intermediate steps where stopping times are needed were glossed over. It might be worthwhile to revisit this analysis in the future, seeing as there does not seem to be much literature along this direction other than the references here in.  $\square$

<sup>4</sup>For more detail on stopping times see [24, Part 2, Chapter II].

*Remark 3.* The definition given for uniformly bounded in probability in this work is stronger than any other definitions that could be found in the literature and follows the classic definition, see Definition 2 and compare it to Definition 11. A trajectory is uniformly bounded in probability, if for all  $r > 0$  there exists a  $B(r) > 0$  such that  $\|x(t_0)\| \leq r$  implies

$$\mathbf{P} \left( \sup_t \|x(t)\| \leq B(r) \right) = 1$$

for all  $t \geq t_0$ . In [21, Definition 2.2] the following definition is given, for all  $r > 0$  and any  $\epsilon > 0$  there exists a  $B(r, \epsilon) > 0$  such that  $\|x(t_0)\| \leq r$  implies

$$\mathbf{P} \left( \sup_t \|x(t)\| \leq B(r) \right) = 1 - \epsilon.$$

In [50, §1.4] the definition of uniformly bounded is defined as follows,

$$\sup_t \mathbf{P}(\|x(t)\| > R) \rightarrow 0 \text{ as } R \rightarrow \infty.$$

Our definition of asymptotic attractivity follows that of Deng and Kristić [21] as it is already as strong as the classic definition. Again however the often cited work by Khasminskii [50] gives a much weaker definition. Compare the following definitions for asymptotically attracting

$$\mathbf{P} \left( \lim_{t \rightarrow \infty} \|x(t)\| = 0 \right) = 1 \quad ([21, \text{Definition 2.2}])$$

$$\lim_{x_0 \rightarrow 0} \mathbf{P} \left( \lim_{t \rightarrow \infty} \|x(t)\| = 0 \right) = 1. \quad ([50, \text{Equation (5.15)}])$$

It is imperative when discussing stability that uniform bounds are achieved.

#### 4.5 Diagonal Stability of Random Matrices

**Theorem 17.** *If  $A_n \in \mathbb{R}^{n \times n}$  is chosen as*

$$[A_n]_{ij} \sim \frac{1}{\sqrt{(2+\delta)n}} \mathcal{N}(0, 1), \quad i \neq j$$

*for any  $\delta > 0$ , and  $[A_n]_{ii} = -1$ , then  $A_n$  is asymptotically almost surely diagonally stable.*

*Proof.* Consider the random matrix  $\bar{A}_n \in \mathbb{R}^{n \times n}$  defined as

$$[\bar{A}_n]_{ij} \sim \frac{1}{\sqrt{(2+\delta)n}} \mathcal{N}(0, 1), \quad i \neq j,$$

and  $[\bar{A}_n]_{ii} = 0$ . Then it follows from (T2) that  $\mathbf{Var}[\bar{A}_n]_{ij} = \frac{1}{(2+\delta)n}$  for  $i \neq j$ . Let  $\bar{B}_n = \bar{A}_n + \bar{A}_n^\top$ , then it follows that  $\mathbf{Var}[\bar{B}_n]_{ij} = \frac{2}{(2+\delta)n}$  when  $i \neq j$  and 0 otherwise. From Theorem 3 we have that

$$\begin{aligned} \sup_{1 \leq i \leq n} \lambda_i(\bar{B}_n) &= 2\sqrt{\frac{2n}{(2+\delta)n}}(1 + o(1)) \\ &< 2(1 + o(1)) \end{aligned}$$

as  $n \rightarrow \infty$  asymptotically almost surely. Noting that  $A_n^\top + A_n \equiv \bar{B}_n - 2I_{n \times n}$  it follows that  $A_n^\top + A_n < 0$  asymptotically almost surely.  $\square$

**Corollary 18.** *If  $A_n \in \mathbb{R}^{n \times n}$  is chosen as*

$$[A_n]_{ij} \sim \mathcal{N} \left( 0, \frac{1}{(2+\delta)n} \right), \quad i \neq j,$$

*for any  $\delta > 0$ , and  $[A_n]_{ii} = -1$ , then  $A_n$  is asymptotically almost surely diagonally stable.*

## 5 Clustering Analysis and Ordination Methods

In the following section clustering techniques are explored in detail as well as ordination techniques for visualizing data.

### 5.1 Distance and Metrics

Let  $x = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n$ , then the Euclidian norm is defined as

$$\|x\| = \left( \sum_{i=1}^n x_i^2 \right)^{1/2}.$$

For a row vector the euclidian norm is similary defined. If no subscript is given with  $\|\cdot\|$  then we assume the Euclidian norm is being used. The following is the Euclidian distance function  $d(x, y) = \|x - y\|$ .

A common distance metric used in ecology is the *Jensen-Shannon Distance* (JSD) metric [46, 54, 59]

$$\text{JSD}(y, z) \triangleq \sqrt{\frac{1}{2} \text{JS}(y, z)}, \quad (\text{T40})$$

where the Jensen-Shanon divergence,  $\text{JS}(y, z) \triangleq \text{KL}(y, \frac{1}{2}(y + z)) + \text{KL}(z, \frac{1}{2}(z + y))$ , is simply the symetrized version of the Kulback-Liebler directed divergence  $\text{KL}(y, z) \triangleq \sum_{i=1}^p y_i \log \frac{y_i}{z_i}$ . So as to not divide by zero, a pseudo count of 1e-10 is added to zero elements before performing the JSD.

### 5.2 $k$ -Medoids

Assume that one has a collection of samples  $X \in \mathbb{R}^{n \times p}$  where  $n$  is the total number of samples and  $p$  is the dimension of each sample. Let  $X_i \in \mathbb{R}^{1 \times p}$ ,  $i = 1, 2, \dots, n$ , be the row vectors of  $X$  as defined below

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}. \quad (\text{T41})$$

We would like each  $X_j$  to belong to a unique cluster within a collection of  $k$  possible clusters  $C_1, C_2, \dots, C_k$ . The unique cluster that contains sample  $X_j$  is denoted  $\mathcal{C}(j; k)$ . If sample  $X_j$  is contained in cluster 3 then  $\mathcal{C}(j) = C_3$ . If one is given an a-priori number of clusters then it is possible to perform this task using the popular paradigm of  $k$ -medoids. The paradigm works as follows. Initially,  $k$  samples are chosen at random as representative *medoids*.  $k$  clusters are then constructed by associating other samples to the nearest medoid. Within each cluster all elements are tested so as to see if a different sample has a smaller within cluster sum of distances. The element with the smallest within cluster sum of distances is chosen as the new medoid for that cluster. This process is performed for each cluster. New clusters are constructed with the  $k$  new medoids and the algorithm repeats again. The MATLAB command `kmedoids` performs  $k$ -medoids with a variety of different methods depending on the number of samples. When the number of samples is less than 3000 MATLAB implements  $k$ -medoids using the *Partitioning Around Medoids* (PAM) algorithm [49].

### 5.3 Silhouette Value

For each sample  $X_j$  there is a corresponding silhouette value  $s_j \in [-1, 1]$  which is defined as follows

$$s_j(k) \triangleq \frac{b_j(k) - a_j(k)}{\max\{a_j(k), b_j(k)\}}$$

where  $a_j$  is the average dissimilarity between sample  $j$  and all other samples within its own cluster,  $b_j$  is the average dissimilarity between  $X_j$  and the elements of the nearest

cluster, and  $k$  is the total number of apriori designated clusters. These two quantities are now formally defined as follows

$$a_j(k) \triangleq \frac{1}{|\mathcal{C}(j; k)| - 1} \sum_{\substack{X_i \in \mathcal{C}(j; k) \\ i \neq j}} d(X_i, X_j).$$

Note that we have used  $|\mathcal{C}(j)|$  to denote the cardinality of  $\mathcal{C}(j)$ , the total number of samples in  $\mathcal{C}(j)$ . We similarly define the  $b_j$  as

$$b_j(k) \triangleq \min_{\mathcal{C}_m \neq \mathcal{C}(j; k)} \frac{1}{|\mathcal{C}_m|} \sum_{X_i \in \mathcal{C}_m} d(X_i, X_j).$$

For a given sample set and a given number of clusters  $k \geq 2$  there is a corresponding  $s_j(k)$ . The Silhouette Index for a sample data set is then the maximum of the mean silhouette value for each total number of clusters.

$$\text{SI}(X) \triangleq \max_k \frac{1}{n} \sum_{j=1}^n s_j(k)$$

The optimal number of clusters is then  $\arg \max_k \frac{1}{n} \sum_{j=1}^n s_j(k)$ . Silhouette analysis is performed in MATLAB using the command `silhouette`.

#### 5.4 Variance Ratio Criterion and the Caliński-Harabasz Index

We now define the *Variance Ratio Criterion* (VRC) which holds for any distance function. When the Euclidian metric is used the VRC is referred to as the *Caliński-Harabasz* (CH) index [12]. As before, assume that a collection of samples has already been grouped into  $k$  clusters. The VRC is defined as

$$\text{VRC}(k) \triangleq \frac{\text{BG}(k)}{\text{WG}(k)} \frac{n - k}{k - 1}$$

where BG is the *Between Group* variance and WG is the *Within Group* variance defined below,

$$\begin{aligned} \text{BG}(k) &\triangleq \sum_{j=1}^k \frac{1}{|\mathcal{C}_j| \sum_{m>j}^k |\mathcal{C}_m|} \sum_{X_i \in \mathcal{C}_j} \sum_{\substack{X_\ell \in \mathcal{C}_m \\ m>j}} d(X_i, X_\ell)^2 \\ \text{WG}(k) &\triangleq \sum_{j=1}^k \frac{2}{|\mathcal{C}_j| (|\mathcal{C}_j| - 1)} \sum_{\substack{i=1 \\ X_i \in \mathcal{C}_j}}^n \sum_{\substack{\ell>i \\ X_\ell \in \mathcal{C}_j}}^n d(X_i, X_\ell)^2. \end{aligned}$$

#### 5.5 Principle Coordinate Analysis

The purpose of *Principle Coordinates Analysis* (PCoA) is to represent a collection of high dimensional data in a lower dimension. Once again assume that one has a collection of samples  $X \in \mathbb{R}^{n \times p}$  where  $n$  is total number of samples and  $p$  is the dimension of each sample. Let  $X_i \in \mathbb{R}^{1 \times p}$ ,  $i = 1, 2, \dots, n$  be the row vectors of  $X$  as defined in (T41). The question answered in this section is how one obtains a  $Y \in \mathbb{R}^{n \times k}$ ,  $k \leq n$  with  $Y_i \in \mathbb{R}^{1 \times k}$ ,  $i = 1, 2, \dots, n$  defined as follows

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

that is a faithful representation of  $X$ . We begin by defining the dissimilarity between samples  $i$  and  $j$  as  $\delta_{ij} = d(X_i, X_j)$ . Then the goal of this method is to find  $Y$  such that  $d(Y_i, Y_j)$  is similar to  $d(X_i, X_j)$  for the distance measure of interest. Let  $D = -\frac{1}{2}(\delta_{ij}^2)_{1 \leq i, j \leq n}$  and

$$B = (I_{n \times n} - n^{-1} \mathbf{1}_n \mathbf{1}_n^T) D (I_{n \times n} - n^{-1} \mathbf{1}_n \mathbf{1}_n^T).$$



where  $\mathbf{1}_n$  is an  $n$ -dimensional column vector with each entry equal to 1. The  $n \times k$  dimensional representation of the  $n \times p$  sample data is then [82, Chapter 5]

$$Y = [q_1\sqrt{\lambda_1}, q_2\sqrt{\lambda_2}, \dots, q_k\sqrt{\lambda_k}].$$

where  $B = Q\Lambda Q^{-1}$  is the eigenvalue decomposition with eigenvalues  $\lambda_i \in \mathbb{R}$ , and normalized eigenvectors  $q_i \in \mathbb{R}^n$  for  $i = 1, 2, \dots, n$  with  $Q = [q_1, q_2, \dots, q_n]$  and  $[\Lambda]_{ii} = \lambda_i$  the diagonal eigenvalue matrix. Due to the fact that  $B$  is symmetric all eigenvalues and eigenvectors will be real valued. It is furthermore assumed that the eigenvalues are arranged such that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . When the Euclidian distance function is used  $B$  is always positive semi-definite. However, it may be possible to have negative eigenvalues when other distance functions are used and thus it is necessary to check and make sure that that  $\lambda_i$  for each  $i \in \{1, 2, \dots, k\}$  is positive for the  $k$ -dimensional space of interest. The command that performs this task in MATLAB is `cmdscale`. MATLAB also rotates the data, but that has no affect on the dissimilarity measure, as rotation is distance preserving in Euclidian space.

## 5.6 Principle Component Analysis

*Principle Component Analysis* (PCA) is similar in spirit to PCoA in that one wishes to represent a set of data in a lower dimension. Assume that one has a collection of samples  $X \in \mathbb{R}^{n \times p}$  where  $n$  is total number of samples and  $p$  is the dimension of each sample. As before the goal is to find a  $Y \in \mathbb{R}^{n \times k}$ ,  $k \leq p$  that is representative of the original data. Assume without loss of generality that the columns of  $X$  have mean zero. Let

$$X = U\Sigma W^T$$

be the singular value decomposition of  $X$  where  $\Sigma \in \mathbb{R}^{n \times p}$  is a diagonal matrix with the singular values  $\sigma_i = [\Sigma]_{ii}$  where  $\sigma_i \geq \sigma_j$  for  $i < j$ . The columns of  $U \in \mathbb{R}^{n \times n}$  are the left singular vectors and the columns of  $W \in \mathbb{R}^{p \times p}$  are the right singular vectors. The reduced order representation is then defined as

$$Y = XW_{1:n,1:k}$$

where  $W_{1:n,1:k}$  contains only the first  $k$  columns of  $W$ . Note that  $W_{1:n,1:k}$  can be thought of as a right hand side projection operator and can be used to project subsequent measurements into pre-existing principle components. Also, note that when the Euclidean distance is used in PCoA it is equivalent to PCA. The command that performs this task in MATLAB is `pca`.

## 6 Modeling

Two approaches for modeling individual human microbial communities are now presented. The first model is the same as that presented in the main text and from this point forward is referred to as the *universal model*. This model assumes that all species interact in a universal manner independent of the host. The second model allows for there to be multiple possible interaction strengths and growth rates for species. This model will be referred to as the *multiple set model*. We use Figure T4 as a visual reference when discussing the modeling paradigms.

### 6.1 Universal Model

Consider a universal species pool indexed by a set of integers  $\mathbf{S} = \{1, \dots, n\}$ , a global  $n \times n$  matrix  $\mathbf{A}$  representing all possible pairwise interactions between species, and a universal vector  $\mathbf{r}$  of size  $n$  containing the growth rates for all the  $n$  species. The global variables for our ecological system are completely defined by the triple  $(\mathbf{S}, \mathbf{A}, \mathbf{r})$ . Then  $q$  *Local Communities* (LCs) are defined by sets  $S^{[\nu]}$ , which are subsets of  $\mathbf{S}$  denoting the specific microbes present in  $LC_\nu$ ,  $\nu = 1, \dots, q$ . For simplicity we assume that each LC contains only  $p$  species ( $p \leq n$ ), randomly selected from the universal species pool. The GLV dynamics for each LC is given by

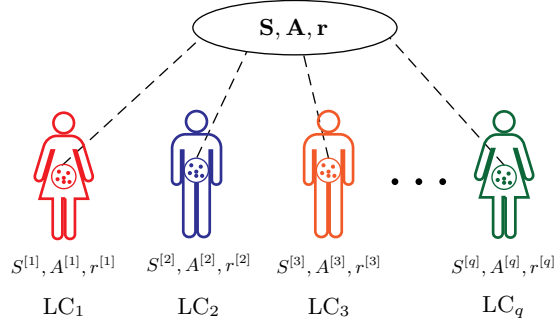
$$LC_\nu : \quad \dot{x}^{[\nu]}(t) = \text{diag}\left(x^{[\nu]}(t)\right) \left(r^{[\nu]} + A^{[\nu]}x^{[\nu]}(t)\right) \quad (\text{T42})$$

where the LC specific interaction matrix and growth vector are defined as  $A^{[\nu]} \triangleq \mathbf{A}_{S^{[\nu]}, S^{[\nu]}}$  and  $r^{[\nu]} \triangleq \mathbf{r}_{S^{[\nu]}}$ , respectively. That is,  $A^{[\nu]}$  is obtained from  $\mathbf{A}$  by only taking the rows and columns of  $\mathbf{A}$  that are contained in the set  $S^{[\nu]}$ . A similar procedure is performed in order to obtain  $r^{[\nu]}$ . Finally there is a map  $m_\nu$  that takes the abundances  $x^{[\nu]}$  in the index  $S^{[\nu]}$  and carries them to the universal index in  $\mathbf{S}$  giving the vector  $\mathbf{x}^{[\nu]} = m_\nu(x^{[\nu]})$  which results in  $\mathbf{x}_{S_j^{[\nu]}}^{[\nu]} = x_j^{[\nu]}$  for  $j = 1, \dots, p$  and  $\mathbf{x}_j^{[\nu]} = 0$  for  $j \in \mathbf{S} \setminus S^{[\nu]}$ . This modeling procedure is inspired by [18]. A toy example of how this model is used to construct a single local community is now given

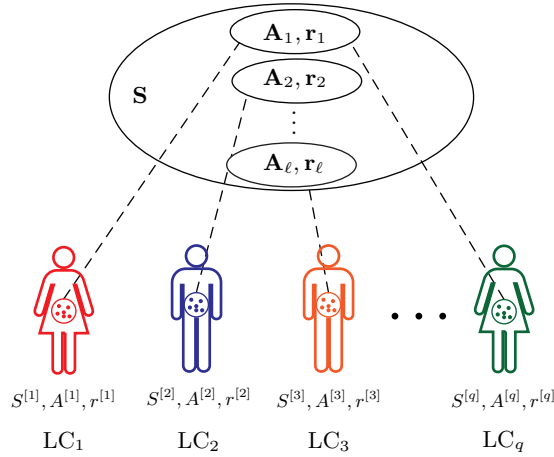
**Example 1.** Consider a global set of 4 species and thus  $\mathbf{S} = \{1, 2, 3, 4\}$ . Let

$$\mathbf{A} = \begin{bmatrix} -1 & 0.1 & 0.4 & -0.1 \\ 0.7 & -1 & 0 & 0.4 \\ -0.1 & 0.7 & -1 & 0 \\ -0.8 & -0.2 & 0.4 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} 0.2 \\ 0.4 \\ 0.5 \\ 0.4 \end{bmatrix}.$$

#### Universal Dynamics



#### Multiple Sets of Dynamics



**Supplementary Text Figure T4: Modeling Paradigms.** Two different modeling paradigms are presented. In the *universal* paradigm there is a universal list of species  $\mathbf{S}$ , a universal interaction matrix  $\mathbf{A}$ , containing all the pairwise interaction strengths, and a universal growth rate vector  $\mathbf{r}$ . Then the dynamics of each Local Community LC is determined by the collection of species in that community. In the *multiple set* paradigm the dynamics are not only determined by the collection of species present but the dynamics are not universal and come from a collection of possible dynamics. In the example above LC<sub>1</sub> dynamics are derived from the first dynamic pair  $(\mathbf{A}_1, \mathbf{r}_1)$ , LC<sub>2</sub> dynamics are derived from  $(\mathbf{A}_2, \mathbf{r}_2)$ , LC<sub>3</sub> dynamics are derived from  $(\mathbf{A}_\ell, \mathbf{r}_\ell)$ , and LC<sub>q</sub> dynamics are derived from the dynamic pair  $(\mathbf{A}_1, \mathbf{r}_1)$ .

Then, if the first local community has the species list  $S^{[1]} = \{1, 3, 4\}$  then it follows that

$$A^{[1]} = \begin{bmatrix} -1 & 0.4 & -0.1 \\ -0.1 & -1 & 0 \\ -0.8 & 0.4 & -1 \end{bmatrix} \quad \text{and} \quad r^{[1]} = \begin{bmatrix} 0.2 \\ 0.5 \\ 0.4 \end{bmatrix}.$$

## 6.2 Multiple Set Model

As before there is a universal set of species  $\mathbf{S} = \{1, \dots, n\}$ , but now there are  $\ell$  pairs of possible global dynamics  $\{(\mathbf{A}_1, \mathbf{r}_1), (\mathbf{A}_2, \mathbf{r}_2), \dots, (\mathbf{A}_\ell, \mathbf{r}_\ell)\}$ . The  $q$  LCs are defined by sets  $S^{[\nu]}$ ,  $\nu = 1, \dots, q$  and a map  $w : \{1, 2, \dots, q\} \rightarrow \{1, 2, \dots, \ell\}$  which determines which model pair the LC is derived from. The GLV dynamics for each LC are given by (T42) where  $A^{[\nu]} \triangleq [\mathbf{A}_{w(\nu)}]_{S^{[\nu]}, S^{[\nu]}}$  and  $r^{[\nu]} \triangleq [\mathbf{r}_{w(\nu)}]_{S^{[\nu]}}$ , respectively. That is,  $A^{[\nu]}$  is obtained from  $\mathbf{A}_{w(\nu)}$  by only taking the rows and columns of  $\mathbf{A}_{w(\nu)}$  that are contained in the set  $S^{[\nu]}$ . A similar procedure is performed in order to obtain  $r^{[\nu]}$ . The map  $m_\nu$  that takes the abundances  $x^{[\nu]}$  in the index  $S^{[\nu]}$  and carries them to the universal index  $\mathbf{S}$  is defined as before  $\mathbf{x}^{[\nu]} = m_\nu(x^{[\nu]})$ . It is worth noting that results from this modeling paradigm are not presented in the main text, because the emergence of community types in this case is trivial (see §7.1) [52].

## 7 Simulation Results and Analysis

We are finally in a position to address the main topic of this work. Revealing properties that allow for community types to arise. Four case studies are now presented. The first case study will use the multiple set model paradigm in Section 6.2, and it will be shown that under this paradigm that clustering of the steady states occurs trivially. The subsequent three case studies use the universal model from Section 6.1. The case studies explore heterogeneity (in terms of interaction strength and network topology), mean degree of the network, and community overlap and how these affect the existence of community types for our dynamics of interest.

### 7.1 Multiple Set Models

For the multiple set model study three universal pairs  $\{(\mathbf{A}_1, \mathbf{r}_1), (\mathbf{A}_2, \mathbf{r}_2), (\mathbf{A}_3, \mathbf{r}_3)\}$  were considered. For this study there was a total of 100 global species. Each  $\mathbf{A}_i$  was a 100 by 100 matrix with values drawn from a normal distribution with mean 0 and variance 0.0049. Then the diagonal elements of each matrix are set to -1, that is  $[\mathbf{A}_i]_{jj} = -1$ ,  $i = 1, 2, 3$ ,  $j = 1, 2, \dots, 100$ . Note that with the above selection, the variance of the off diagonal elements of  $\mathbf{A}_i$  satisfy the following bound  $0.0049 < \frac{1}{2 \cdot 100}$ , which from Corollary 18 is the sufficient condition on the variance for asymptotic almost sure diagonal stability when  $n = 100$ . Also, note the interaction network is implicitly a complete graph without any structural heterogeneity. Finally, each  $\mathbf{r}_i$  had elements drawn from a uniform distribution between 0 and 1.

From the three pairs above 600 local communities were constructed, 200 from each universal pair. Each of the local communities contained 80 species. The dynamics were then simulated for 100 seconds with initial conditions drawn from  $\mathcal{U}(0, 1)$ . The abundances of the species in the communities were then normalized, and the relative abundances of the 600 local communities were clustered using  $k$ -medoids from Section 5.2 and silhouette indexed as defined in Section 5.3, each using the Jensen-Shannon metric in (T40). A principle coordinate plot of the results is given in Figure T5 with the elements colored according to cluster assignment from  $k$ -medoids. It is clear that there are three clusters. Furthermore each cluster exactly coincides with the universal  $\mathbf{A}_i$  that the local community dynamics were obtained from.

### 7.2 Universal Model

For the three universal case studies the global interaction matrix is defined as

$$\mathbf{A} = \mathbf{NH} \circ \mathbf{G}s,$$

which contains four components.

- (i)  $\mathbf{N} \in \mathbb{R}^{n \times n}$  is the nominal component where each element is sampled from either a normal distribution or a uniform distribution.
- (ii) The matrix  $\mathbf{H}$  is a diagonal matrix that captures the overall interaction heterogeneity of different species. When interaction strength heterogeneity is employed the diagonal elements of  $\mathbf{H}$  are drawn from a power-law distribution (T3) with exponent  $-\alpha$ ,  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$ , which are subsequently normalized so that the mean of the diagonal is equal to 1. Without interaction heterogeneity  $\mathbf{H}$  is simply the identity matrix.
- (iii) The matrix  $\mathbf{G}$  is the adjacency matrix of the underlying ecological network:  $[\mathbf{G}]_{ij} = 1$  if species  $i$  is affected by the presence of species  $j$  and 0 otherwise. When the underlying network is a complete digraph all elements in  $\mathbf{G}$  are equal to 1. For details on the construction of  $\mathbf{G}$  when the underlying network is Erdős-Rényi or a power-law digraph see Sections 3.2.1 and 3.2.2 respectively.
- (iv) The last element  $s$  is simply a scaling factor between 0 and 1.

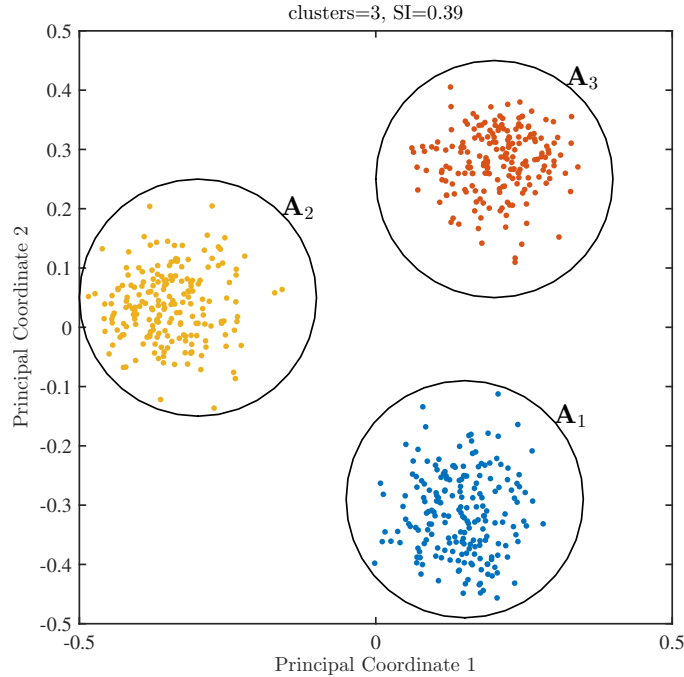
As before we set  $[\mathbf{A}]_{ii} = -1$  to ensure one of the necessary conditions for diagonal stability of a matrix is satisfied (Theorem 10 and Corollary 11). Finally, the elements in the global growth rate vector are defined as

$$\mathbf{r} = \mathbf{h} \circ \mathbf{n}$$

where  $\mathbf{n}$  is the nominal component taken from a uniform distribution, and  $\mathbf{h}$  captures the growth rate heterogeneity. When there is growth rate heterogeneity  $\mathbf{h}$  is drawn from a power-law distribution with exponent  $-\alpha$  and subsequently normalized to have a mean of 1. Without growth rate heterogeneity  $\mathbf{h}$  is simply a column vector of ones. Note that  $\mathbf{h}$  is not included in the main text, and is only used in one scenario of one case study within this section.

### 7.2.1 Universal Model: Heterogeneity Study

Next is a systematic study of heterogeneity and its effects on the dynamics in the universal model. Eight different scenarios were tested in this study. For each of the following scenarios the number of global species is  $n = 100$ . Table T1 outlines the differences between the scenarios.



**Supplementary Text Figure T5:** Principle coordinates for multiple model case study.

- Scenario 1.  $[\mathbf{N}]_{ij} \sim \mathcal{U}(-0.5, 0.5)$ ,  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$  where  $\alpha \in [1.2, 7]$  and subsequently normalized to have a mean of 1,  $\mathbf{G}$  is the adjacency matrix for a complete digraph and thus all entries are equal to 1, and the scaling factor is set to  $s = 0.07$ . There is no growth rate heterogeneity and thus the column vector  $\mathbf{h}$  is all ones and finally  $[\mathbf{n}]_i \sim \mathcal{U}(0, 1)$ .
- Scenario 2. The same as Scenario 1 but with  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$  and  $s = 0.07$ .
- Scenario 3.  $[\mathbf{N}]_{ij} \sim \mathcal{U}(-0.5, 0.5)$ ,  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$  where  $\alpha \in [1.2, 7]$  and subsequently normalized to have a mean of 1,  $\mathbf{G}$  is the adjacency matrix for an Erdős-Rényi digraph with a mean out-degree of 10, and the scaling factor is set to  $s = 0.5$ . There is no growth rate heterogeneity and thus the column vector  $\mathbf{h}$  is all ones and finally  $[\mathbf{n}]_i \sim \mathcal{U}(0, 1)$ . For more details on Erdős-Rényi digraphs see §3.2.1.
- Scenario 4. The same as Scenario 3 but with  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$  and  $s = 0.1$ .
- Scenario 5.  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$ ,  $\mathbf{H}$  is the identity matrix,  $\mathbf{G}$  is the adjacency matrix for a digraph with the out-degree drawn from a power-law distribution  $\mathcal{P}(\alpha)$  with a mean out-degree of 10 where  $\alpha \in [1.2, 7]$ , and the scaling factor is set to  $s = 0.2$ . There is no growth rate heterogeneity and thus the column vector  $\mathbf{h}$  is all ones and  $[\mathbf{n}]_i \sim \mathcal{U}(0, 1)$ . For more details on power-law out-degree digraphs see §3.2.2.
- Scenario 6.  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$ ,  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$  where  $\alpha \in [1.2, 7]$  and subsequently normalized to have a mean of 1,  $\mathbf{G}$  is the adjacency matrix for a digraph with the out-degree drawn from a power-law distribution  $\mathcal{P}(\alpha)$  with a mean out-degree of 10 where  $\alpha \in [1.2, 7]$ , and the scaling factor is set to  $s = 0.1$ . The power-law degree distribution and interaction strength heterogeneity are uncoupled. There is no growth rate heterogeneity and thus the column vector  $\mathbf{h}$  is all ones and  $[\mathbf{n}]_i \sim \mathcal{U}(0, 1)$ . For more details on power-law out-degree digraphs see §3.2.2.
- Scenario 7.  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$ ,  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$  where  $\alpha \in [1.2, 7]$  and subsequently normalized to have a mean of 1,  $\mathbf{G}$  is the adjacency matrix for a digraph with the out-degree drawn from a power-law distribution  $\mathcal{P}(\alpha)$  with a mean out-degree of 10 where  $\alpha \in [1.2, 7]$ , and the scaling factor is set to  $s = 0.02(\alpha + 1)$ . The power-law degree distribution and interaction strength heterogeneity are coupled so that the node with the highest out-degree is also the node with the largest interaction strength scaling. There is no growth rate heterogeneity and thus the column vector  $\mathbf{h}$  is all ones and  $[\mathbf{n}]_i \sim \mathcal{U}(0, 1)$ . For more details on power-law out-degree digraphs see §3.2.2.
- Scenario 8.  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$ ,  $\mathbf{H}$  is the identity matrix,  $\mathbf{G}$  is the adjacency matrix for an Erdős-Rényi digraph with a mean out-degree of 10, and the scaling factor is set to  $s = 0.1$ . There is growth rate heterogeneity and thus the column vector  $\mathbf{h}$  has elements drawn from a power-law distribution  $\mathcal{P}(\alpha)$  and subsequently normalized to have a mean of 1 and  $[\mathbf{n}]_i \sim \mathcal{U}(0, 1)$ . For more details on Erdős-Rényi digraphs see §3.2.1.

For each of the eight scenarios above and for every  $\alpha$ ,  $q = 500$  local communities were generated each with  $p = 80$  species selected at random from the  $n = 100$  global species. The dynamics as described above following the modeling paradigm in Section 6.1 were then simulated for 100 seconds with initial conditions drawn from  $\mathcal{U}(0, 1)$ . If any of the 500 simulations crashed due to instability or if the norm of the terminal discrete time derivative was greater than 0.01 then that local community was excluded from the rest of the study. Those simulations that finished without crashing and with small terminal discrete time derivative were deemed steady. Less than 1% of simulations were deemed unstable. The abundances of the species in the communities were then normalized, and the relative abundances of the 500 local communities were clustered using  $k$ -medoids from

Section 5.2 and silhouette indexed as defined in Section 5.3, each using the Jensen-Shannon distance metric in (T40).

The results from the above scenarios are presented in Figures T6-T13. The first plot is a comprehensive clustering analysis of the steady state values obtained from the simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots

$$\max_{i,j} \operatorname{real} \left( \lambda_i(A^{[j]}) \right)$$

as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .

Figures T6 and T7 show that regardless of whether the nominal component is drawn from a uniform or a normal distribution the increase in interaction strength heterogeneity (decreasing  $\alpha$ ) leads to steady state clustering in the data. Figures T8 and T9 illustrate that the same phenomenon also holds for Erdős-Rényi digraphs as well. When the underlying graph topology follows a power-law degree distribution, but there is no interaction strength heterogeneity, clustering of steady states is not observed, see Figure T10. Figure T11 illustrates the fact that when there is interaction strength heterogeneity and network degree heterogeneity it is possible to have clustering of steady states when  $\alpha$  is in the range  $[3, 1]$ . However the trend is not smooth and is inconsistent, as compared to Figures T6-T9. This is due to the fact that when the underlying interaction network is being constructed, there is no guarantee that the high-degree node will also have a large interaction strength. For instance, if a node with no out edges is randomly selected to have high interaction strength scaling, then the impact of that node on the rest of the nodes is still zero. When the interaction strength heterogeneity is coupled with the out-degree for a power-law out-degree digraph then the trends from Figures T6-T9 are recovered. Finally, if the growth rates of the species are derived from a power-law distribution, then clustering of steady states also occurs as  $\alpha$  decreases.

Rows three and four illustrate the spectrum of  $A^{[j]}$ , and are included so that we can infer the asymptotic stability of the system for certain paradigms. Note that regardless of whether the nominal interactions are drawn from a normal distribution or a uniform distribution when  $\alpha$  is large the spectrum represents a uniform disk in the complex plane as predicted by Theorem 1. Furthermore, for scenario 2 with  $s = 0.07$  and  $[\mathbf{N}]_{ij} \sim \mathcal{N}(0, 1)$  it follows that  $\operatorname{Var}[\mathbf{A}]_{ij} < \frac{1}{\sqrt{2n}}$  for  $n = 100$ . From Theorem 17 it follows that  $\mathbf{A}$  is diagonally stable, in a probabilistic sense. Then invoking Theorem 10 we know that any principle minor of  $\mathbf{A}$  is diagonally stable as well. Therefore, in Scenario 2 for large  $\alpha$  each  $A^{[j]}$  is diagonally stable, in a probabilistic sense.

Also, for all of the scenarios with interaction heterogeneity all of the eigenvalues of  $\mathbf{A}$ , and consequently  $A^{[j]}$ , converge to  $-1$  as  $\alpha$  tends to 1. In the limit of  $\alpha$  tending to 1, only one of the columns of  $\mathbf{A}$  has non-zero values off the diagonal. Therefore, in the limit of  $\alpha$  tending to 1 the following inequality holds  $\mathbf{A}^\top P + P\mathbf{A} < 0$  where  $P = [1, \epsilon, \dots, \epsilon]^\top$  and  $\epsilon$  is sufficiently small.<sup>5</sup> Thus  $\mathbf{A}$  is diagonally stable in the limit of  $\alpha$  tending to 1. Therefore, for low interaction strength heterogeneity and for high interaction strength heterogeneity one of the necessary conditions for uniform asymptotic stability in the positive orthant is satisfied, see Theorem 5. Note that even when the  $A^{[j]}$  are not diagonally stable it does not imply that the state trajectory  $x^{[j]}(t)$  is unstable.

### 7.2.2 Universal Model: Sparsity Study

Next is a study where the mean degree of the Erdős-Rényi digraph along with the interaction strength heterogeneity is varied. Table T2 outlines the differences between the scenarios. The details of the study are as follows:  $[\mathbf{N}]_{ij} \sim \mathcal{U}(-0.5, 0.5)$ ,  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$  where  $\alpha \in [1.2, 7]$  and subsequently normalized to have a mean of 1,  $\mathbf{G}$  is the adjacency

<sup>5</sup>Without loss of generality we have assumed that the first column of  $\mathbf{A}$  is the highly weighted column.

matrix for an Erdős-Rényi digraph with a mean out-degree of

$$d \in \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\},$$

and the scaling factor is set to  $s = 1/\sqrt{d}$ . There is no growth heterogeneity and thus the column vector  $\mathbf{h}$  is all ones and finally  $[\mathbf{n}]_i \sim \mathcal{U}(0, 1)$ . For each of the ten scenarios above and for every  $\alpha$  the same procedure as in §7.2.1 was carried out and the results are shown in Figures T14-T23.

From Figures T14-T23 it can be concluded that so long as the mean out-degree of the ER digraph is greater than 2 the steady states of the GLV model increase in SI as  $\alpha$  decreases. That is, the same trends as observed in the previous study hold, so long as the ER digraph is connected. When the mean degree is 1 for an ER graph it is very unlikely that the graph will be connected [29]. Thus for a digraph with mean out-degree 1, it is even more unlikely that it will be connected. When the underlying digraph  $\mathbf{G}$  has many isolated nodes then the scaling of interaction strengths does not influence as many other species in the GLV system and thus clustering is not observed.

### 7.2.3 Universal Model: Community Size Study

In all previous studies each local community contained 80 species. For this study the size of the local communities take on values in the following set

$$p \in \{100, 99, 95, 90, 80, 70, 60, 50\}.$$

Details of the study are as follows:  $[\mathbf{N}]_{ij} \sim \mathcal{U}(-0.5, 0.5)$ ,  $[\mathbf{H}]_{ii} \sim \mathcal{P}(\alpha)$  where  $\alpha \in [1.2, 7]$  and subsequently normalized to have a mean of 1,  $\mathbf{G}$  is the adjacency matrix for an Erdős-Rényi digraph with a mean out-degree of 10, and the scaling factor is set to  $s = 0.5$ . There is no growth heterogeneity and thus the column vector  $\mathbf{h}$  is all ones and finally  $[\mathbf{n}]_i \sim \mathcal{U}(0, 1)$ . Table T3 outlines the differences between the scenarios. For each of the eight scenarios above and for every  $\alpha$  the same procedure as in §7.2.1 was carried out and the results are shown in Figures T24-T31.

These results show that the trends observed in the earlier studies still hold when the community sizes are varied between 95 and 50 species, Figures T26-T31. As the community sizes approach 50 the trend of increased clustering with increased heterogeneity is less significant. When the number of species in the LCs approaches the number of species in the meta-community, 100, the results do not follow the same trends as before and the Silhouette Indices are near 1, independent of the interaction strength heterogeneity. In Figure T24 all the simulations are identical, each LC has the same 100 species but with different initial conditions. Due to asymptotic stability all of the simulations converge to the same steady state (only those simulations for  $\alpha$  between 2.2 and 3 have the potential to be unstable). Even though the Silhouette Indices are near 1 across the heterogeneity spectrum, all of the steady state values are within  $10^{-3}$  to  $10^{-2}$  in terms of the first two principle coordinates. When all of the LCs differ by only one species, Figure T25, once again the Silhouette Indices are near 1, yet the PCoA illustrates that most of the samples are very close together with just a few outliers. Figures T24 and T25 illustrate the sometimes confounding results when clustering analysis and PCoA are performed [44, 67].

**Table T1:** Parameter settings for heterogeneity case study.

Scenario	Growth Rate Distribution	Nominal Interaction Distribution	Interaction Structure	Mean Degree	Interaction Strength Heterogeneity	Network Structure Heterogeneity	Growth Rate Heterogeneity
1	uniform	uniform	complete	100	$\alpha \in [1.2, 7]$	none	none
2	uniform	normal	complete	100	$\alpha \in [1.2, 7]$	none	none
3	uniform	uniform	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	none
4	uniform	normal	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	none
5	uniform	normal	power-law	10	none	$\alpha \in [1.2, 7]$	none
6	uniform	normal	power-law	10	$\alpha \in [1.2, 7]$	$\alpha \in [1.2, 7]$	none
7	uniform	normal	power-law	10	$\alpha \in [1.2, 7]$	$\alpha \in [1.2, 7]$ (coupled)	none
8	uniform	normal	Erdős-Rényi	100	none	none	$\alpha \in [1.2, 7]$

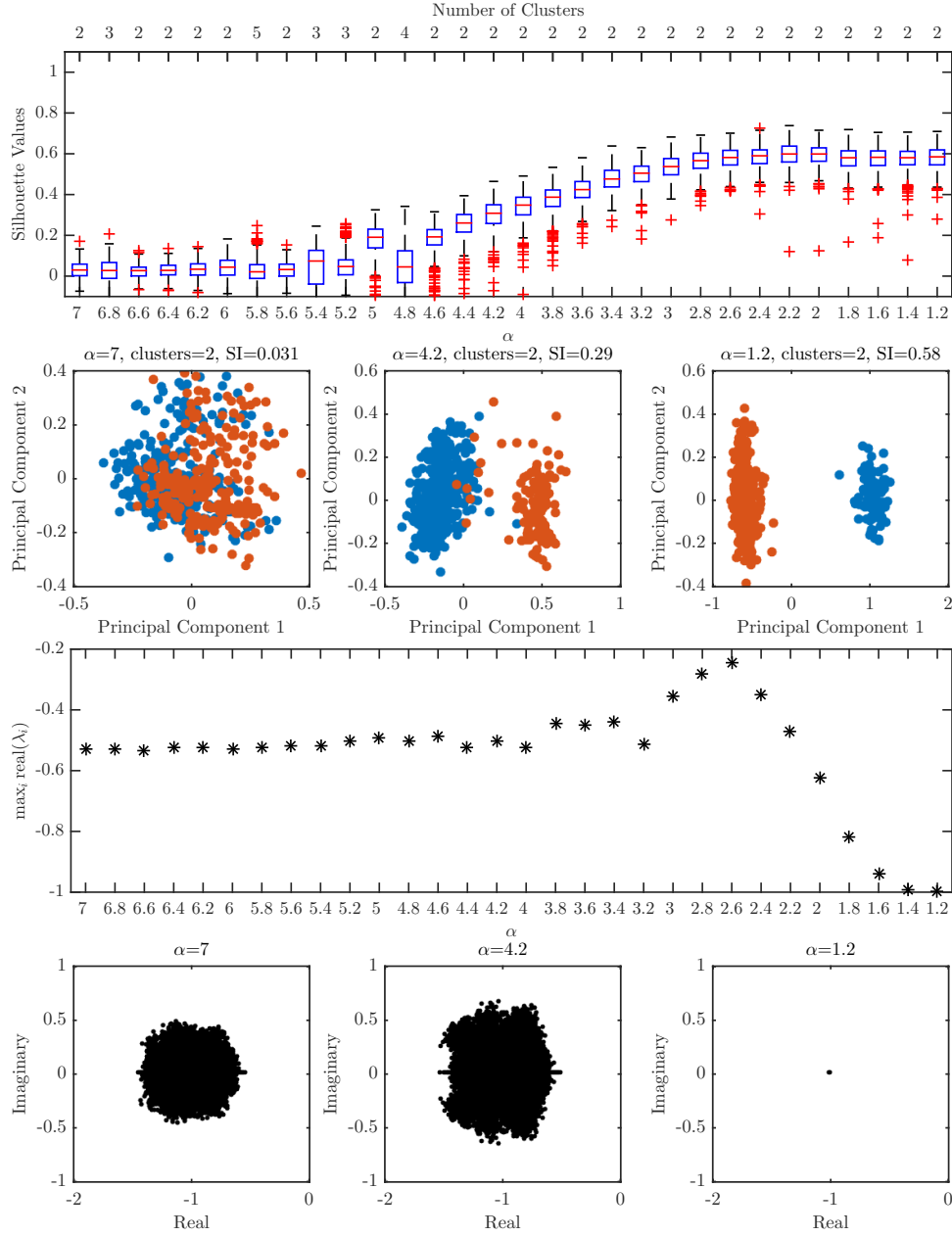


**Table T2:** Parameter settings for sparsity study.

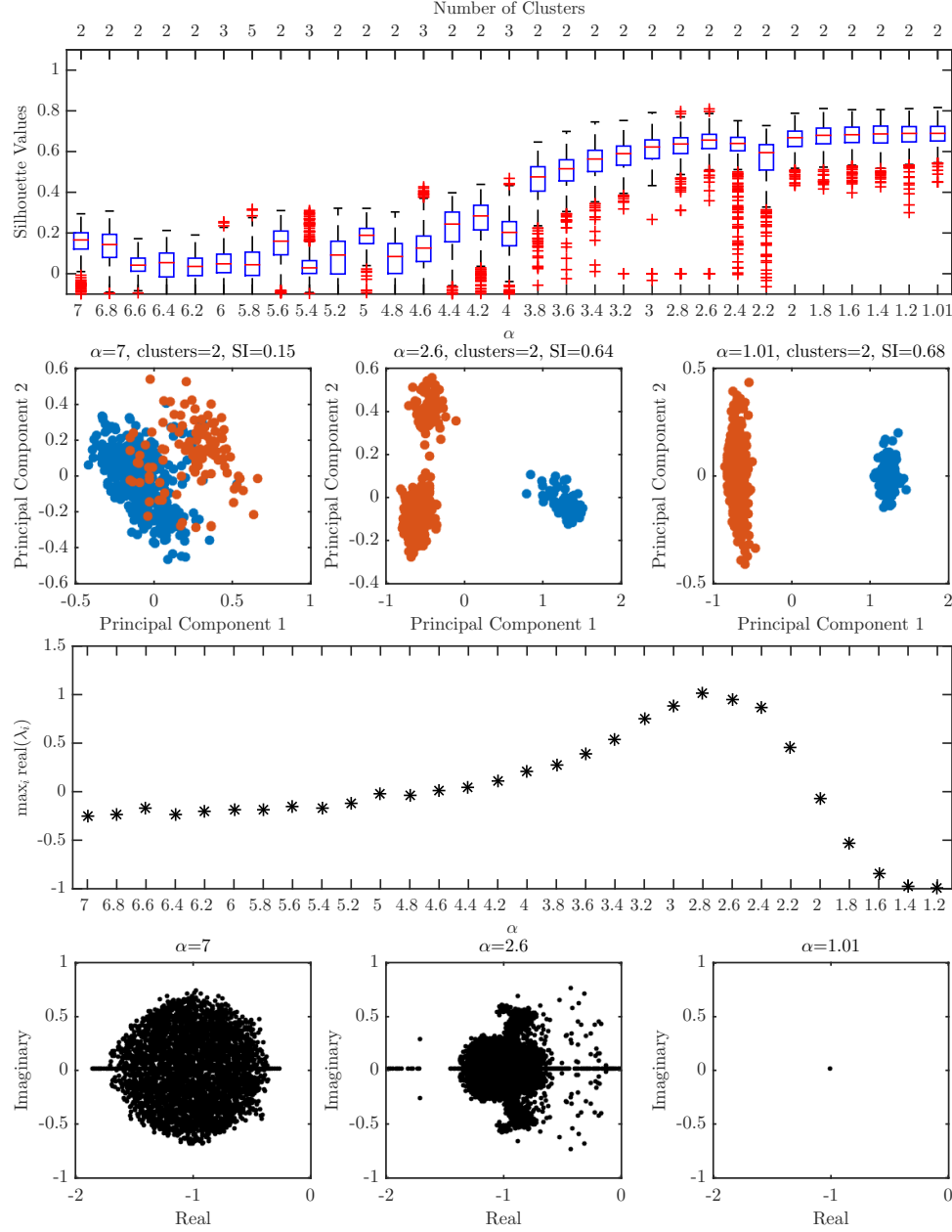
Scenario	Growth Rate Distribution	Nominal Interaction Distribution	Interaction Structure	Mean Degree	Interaction Strength Heterogeneity	Network Structure Heterogeneity	Growth Rate Heterogeneity
1	uniform	uniform	Erdős-Rényi	1	$\alpha \in [1.2, 7]$	none	none
2	uniform	uniform	Erdős-Rényi	3	$\alpha \in [1.2, 7]$	none	none
3	uniform	uniform	Erdős-Rényi	5	$\alpha \in [1.2, 7]$	none	none
4	uniform	uniform	Erdős-Rényi	7	$\alpha \in [1.2, 7]$	none	none
5	uniform	uniform	Erdős-Rényi	9	$\alpha \in [1.2, 7]$	none	none
6	uniform	uniform	Erdős-Rényi	11	$\alpha \in [1.2, 7]$	none	none
7	uniform	uniform	Erdős-Rényi	13	$\alpha \in [1.2, 7]$	none	none
8	uniform	uniform	Erdős-Rényi	15	$\alpha \in [1.2, 7]$	none	none
9	uniform	uniform	Erdős-Rényi	17	$\alpha \in [1.2, 7]$	none	none
10	uniform	uniform	Erdős-Rényi	19	$\alpha \in [1.2, 7]$	none	none

**Table T3:** Parameter settings for community size study.

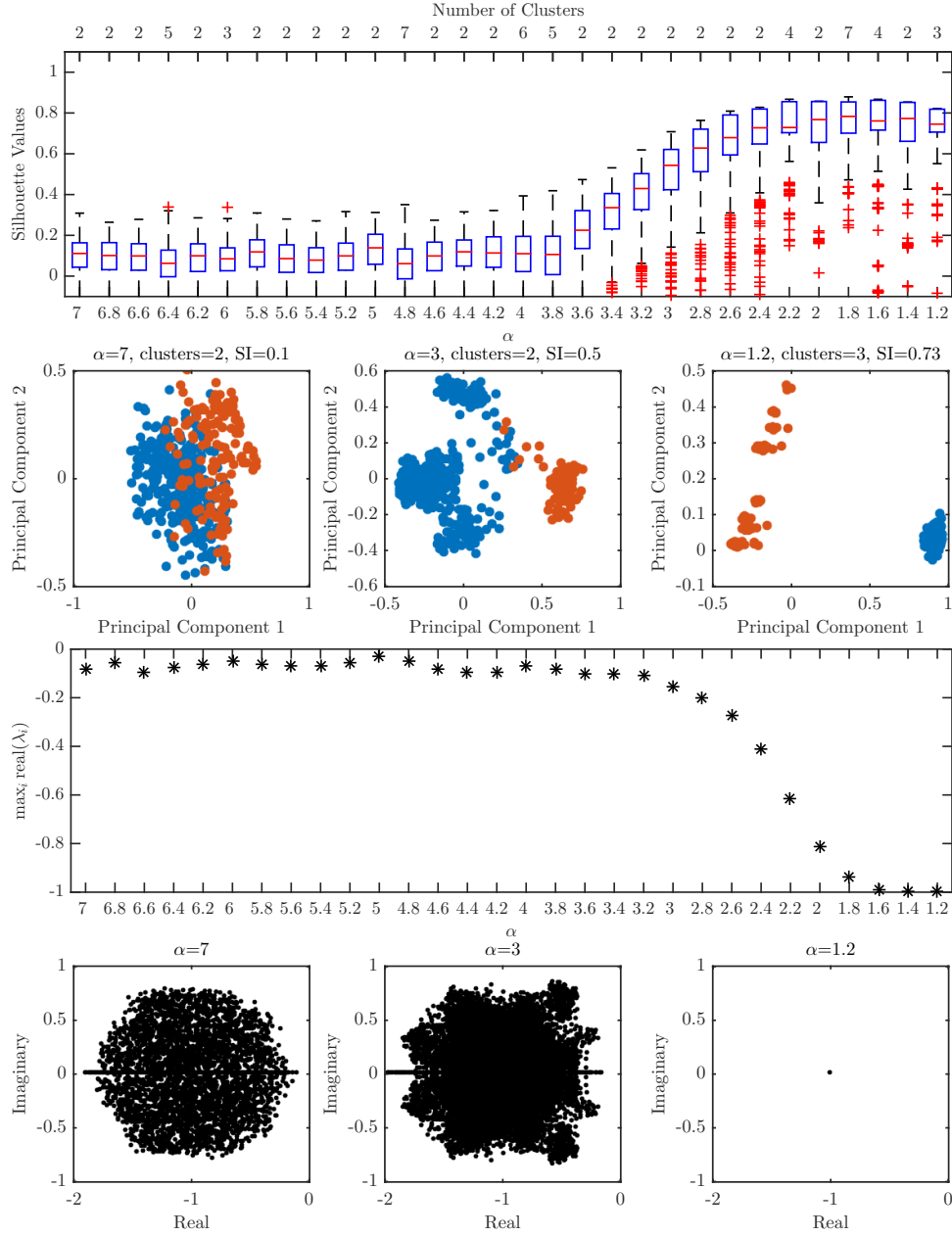
Scenario	Growth Rate Distribution	Nominal Interaction Distribution	Interaction Structure	Mean Degree	Interaction Strength Heterogeneity	Network/Growth Heterogeneity	Community Size
1	uniform	uniform	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	100
2	uniform	uniform	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	99
3	uniform	uniform	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	95
4	uniform	uniform	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	90
5	uniform	uniform	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	80
6	uniform	uniform	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	70
7	uniform	uniform	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	60
8	uniform	uniform	Erdős-Rényi	10	$\alpha \in [1.2, 7]$	none	50



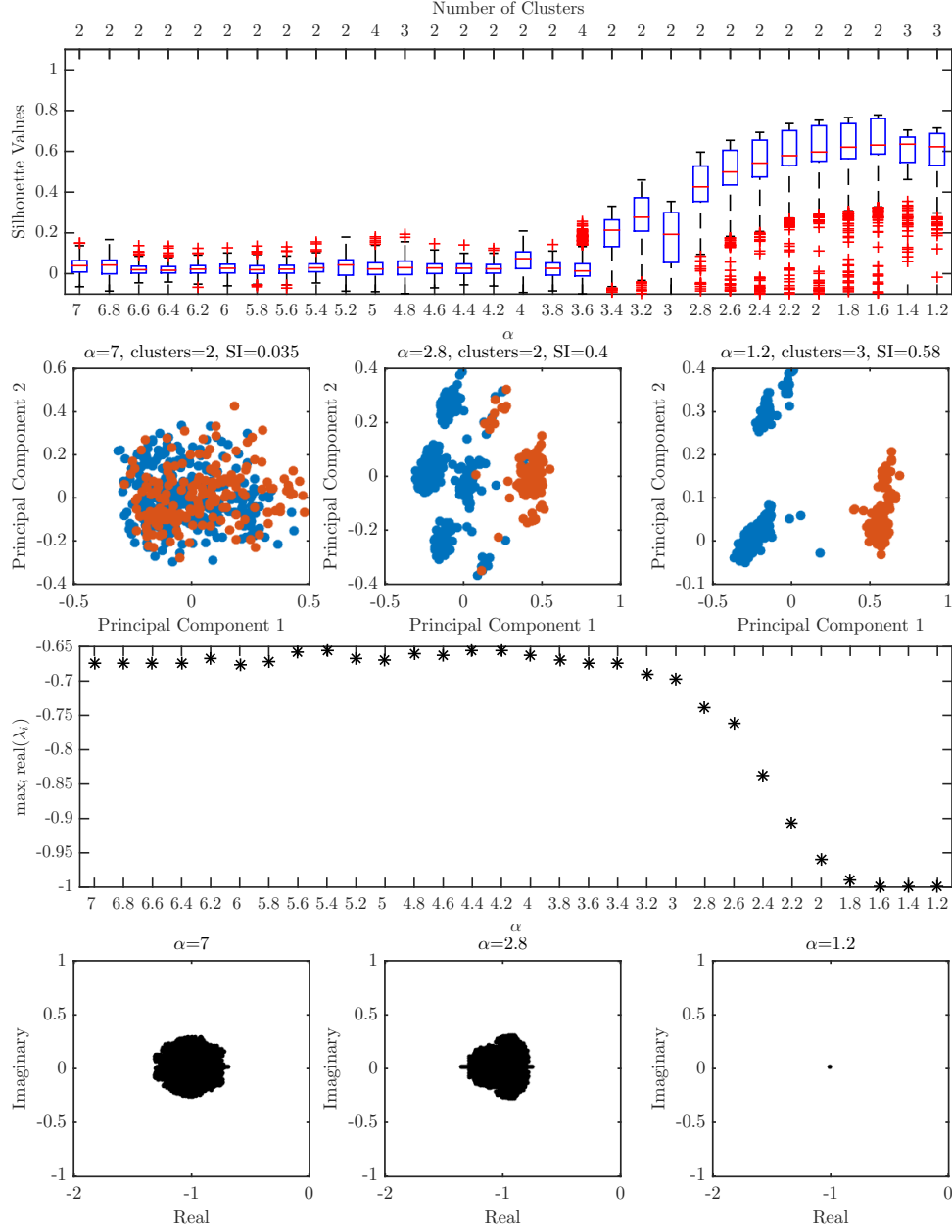
**Supplementary Text Figure T6:** Universal Model Heterogeneity Study Scenario 1 in Table T1. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



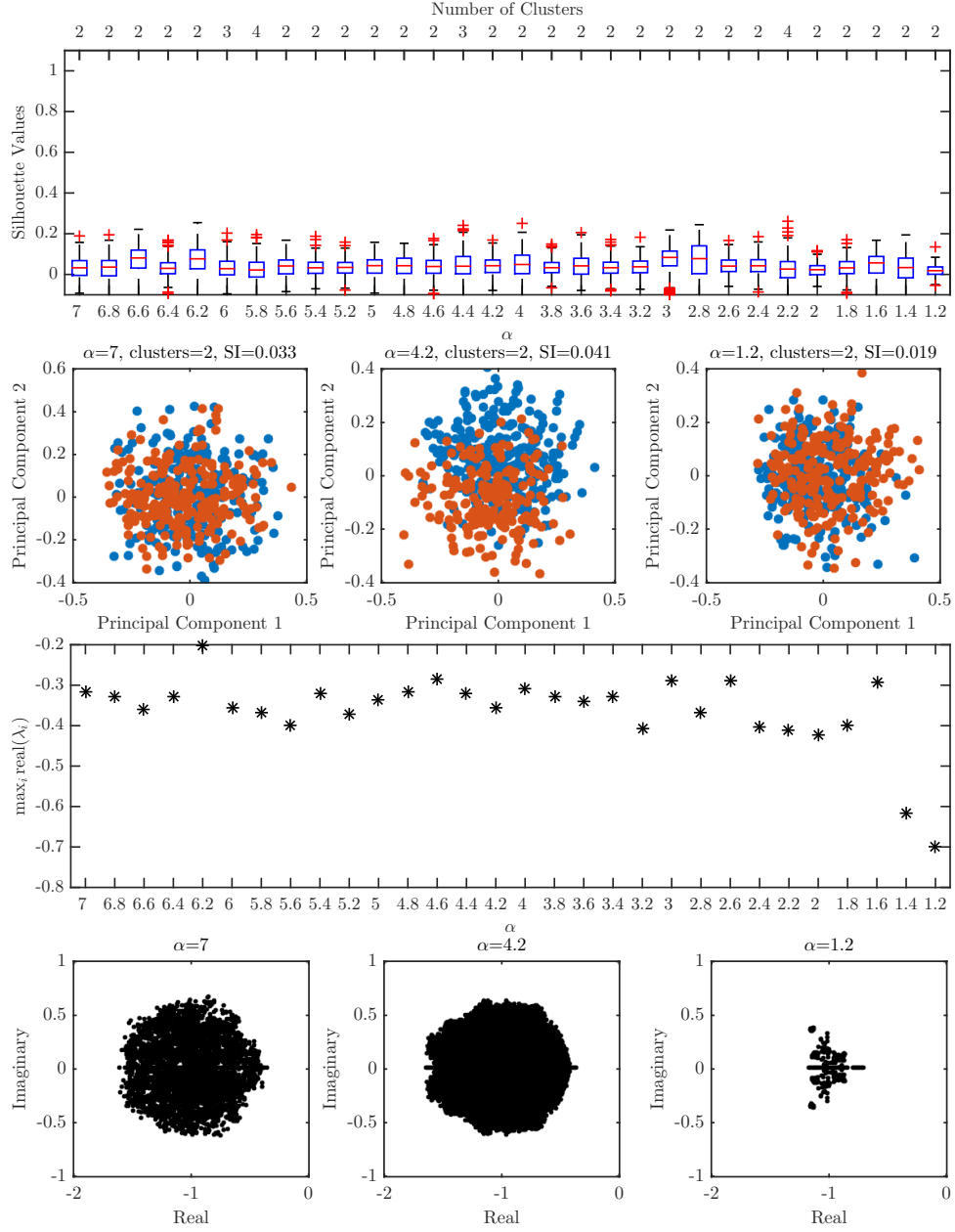
**Supplementary Text Figure T7:** Universal Model Heterogeneity Study Scenario 2 in Table T1. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



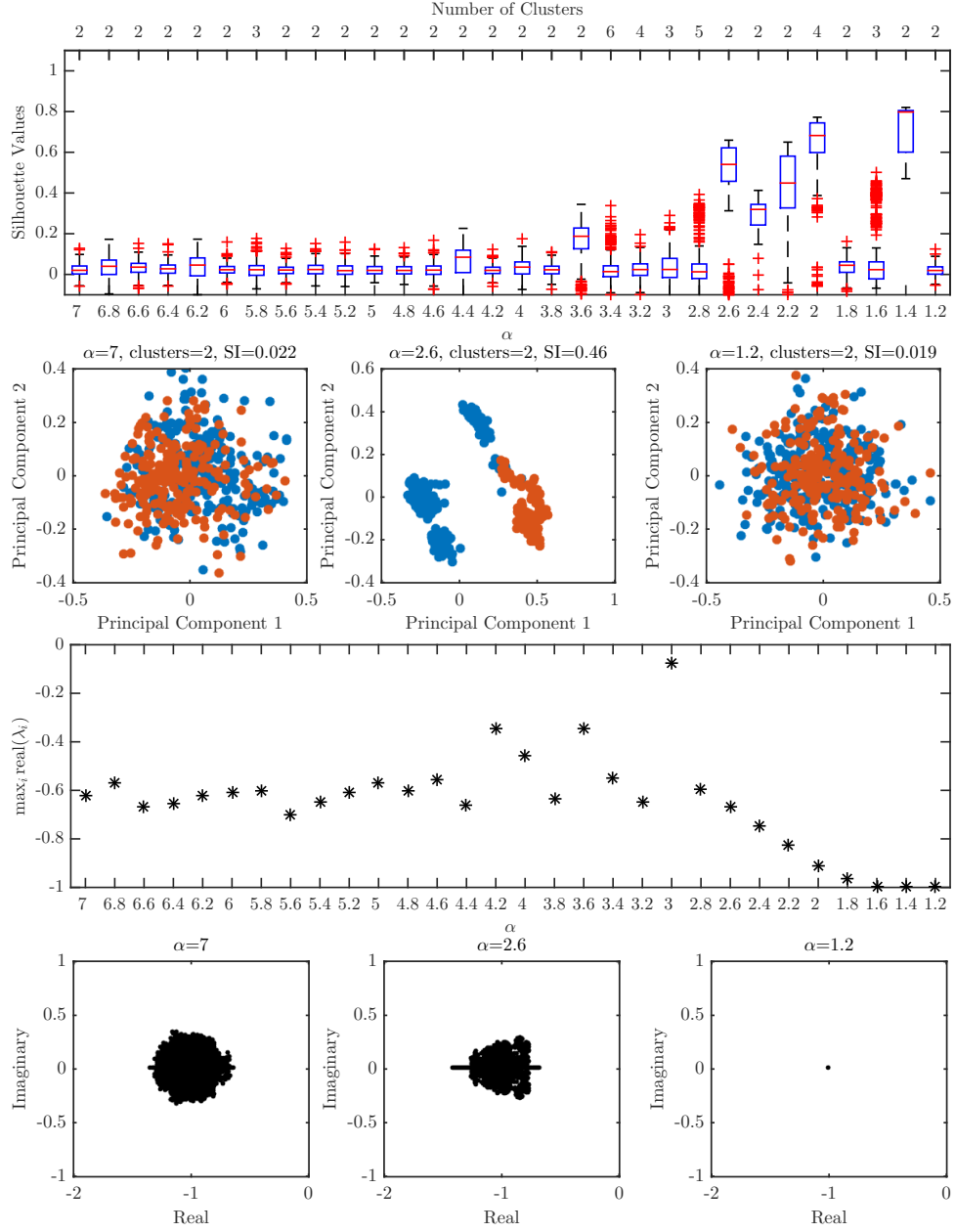
**Supplementary Text Figure T8:** Universal Model Heterogeneity Study Scenario 3 in Table T1. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



**Supplementary Text Figure T9:** Universal Model Heterogeneity Study Scenario 4 in Table T1. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .

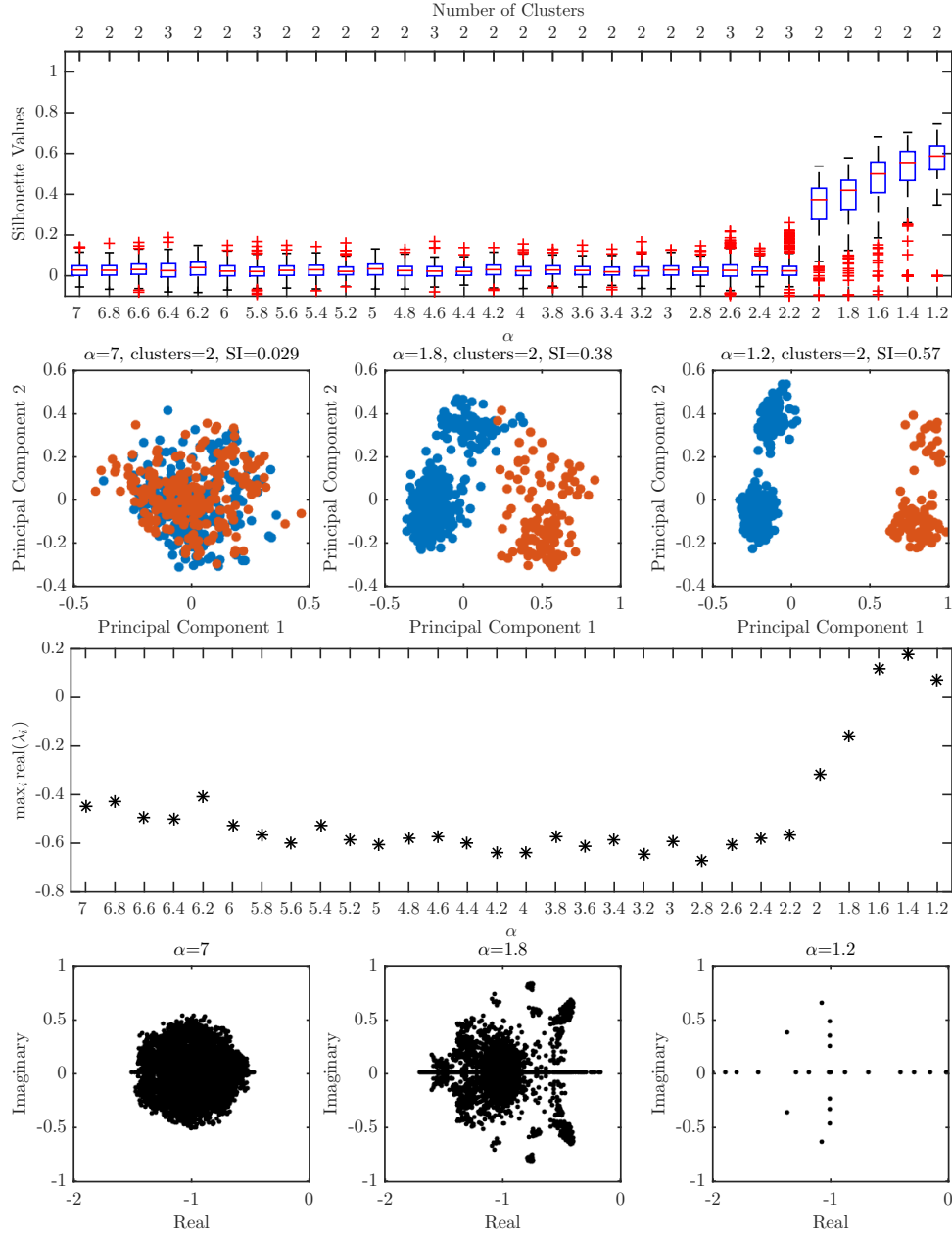


**Supplementary Text Figure T10:** Universal Model Heterogeneity Study Scenario 5 in Table T1. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .

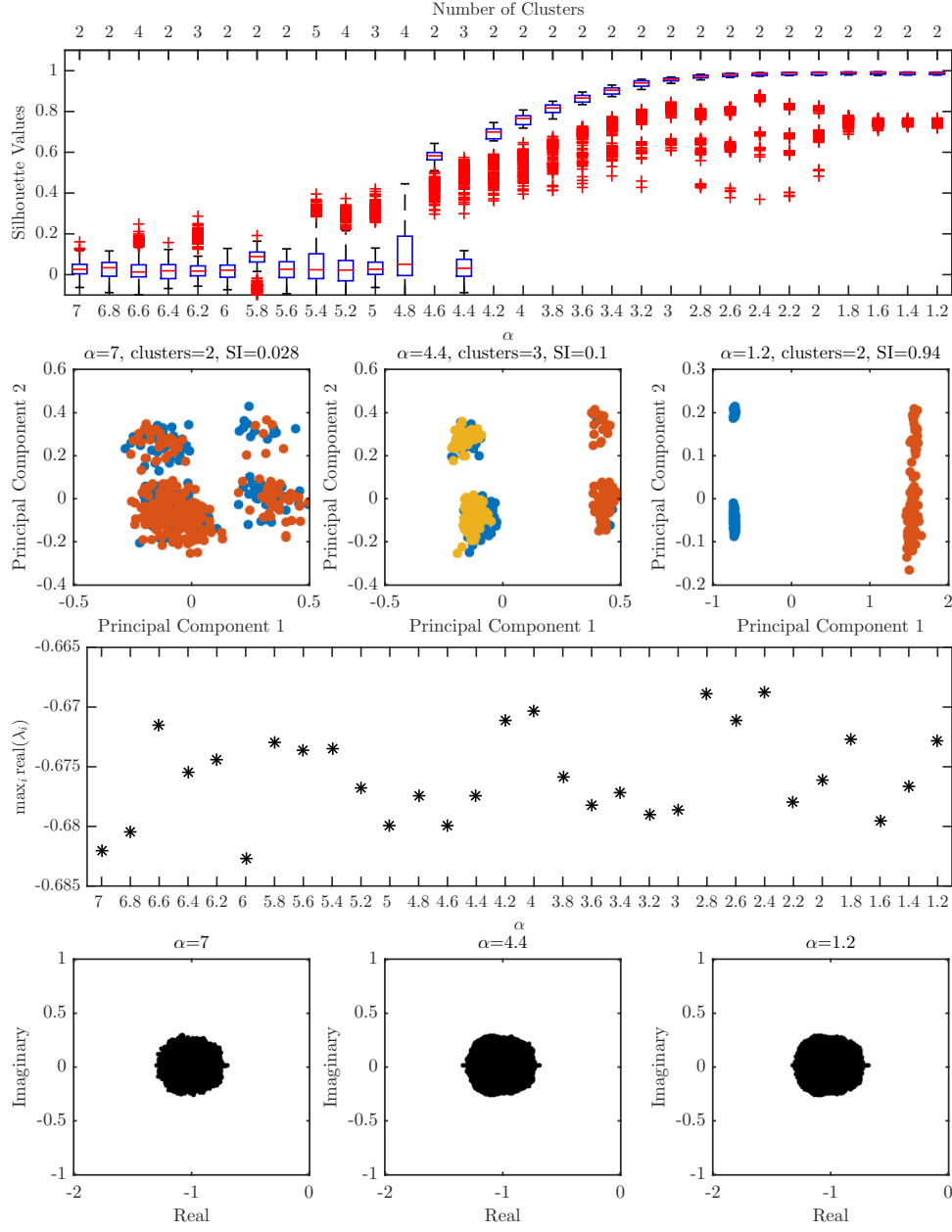


**Supplementary Text Figure T11:** Universal Model Heterogeneity Study Scenario 6 in Table T1. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .

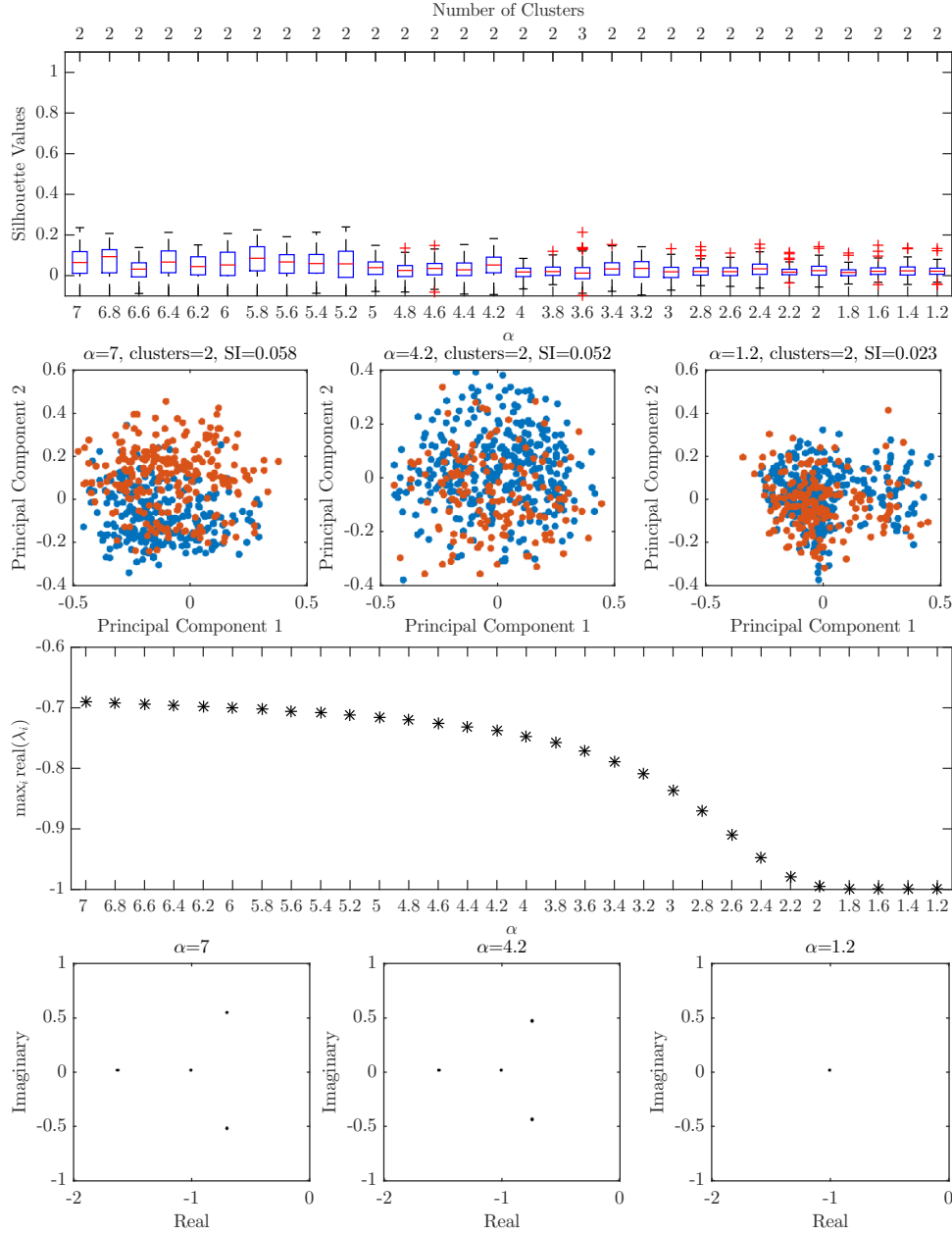




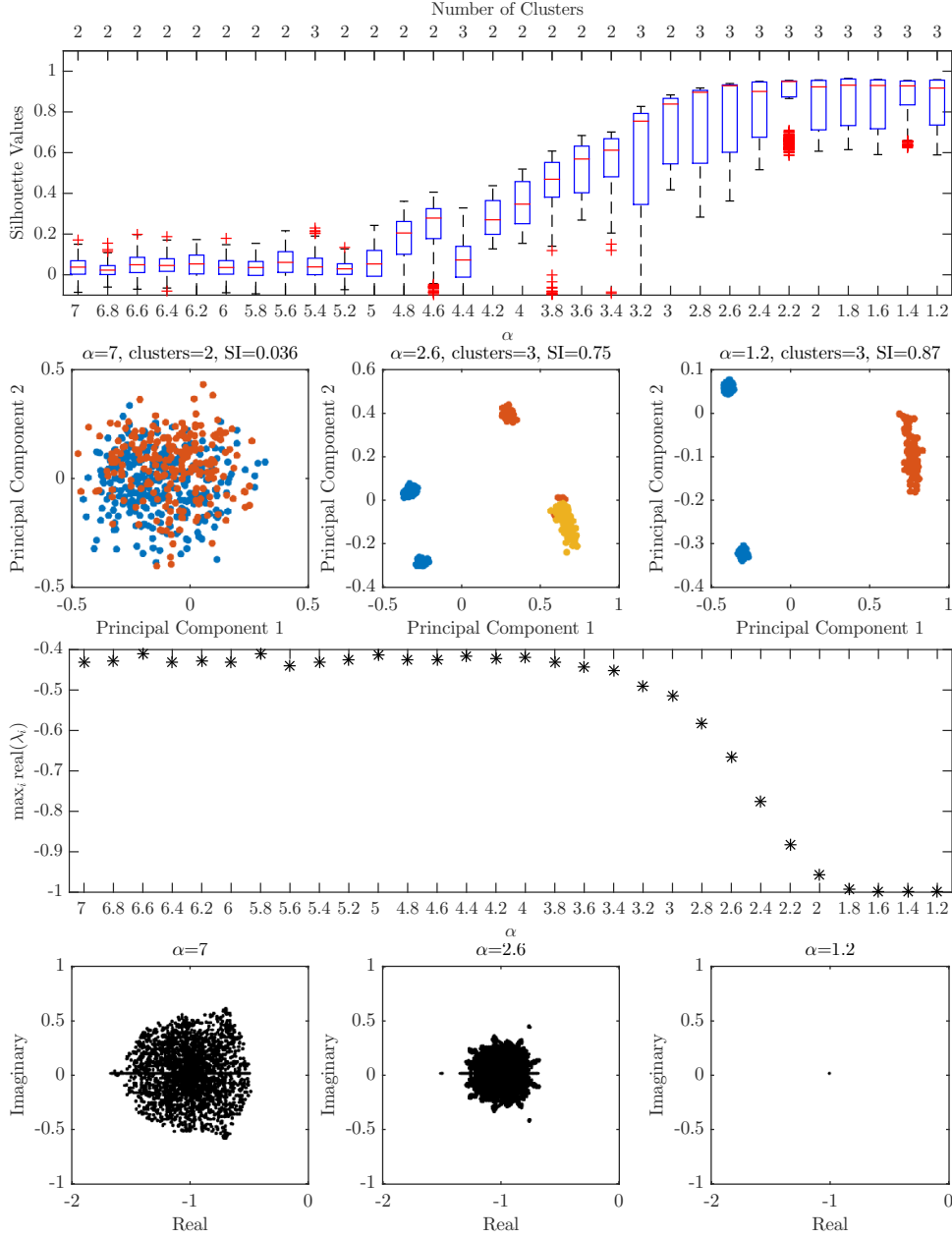
**Supplementary Text Figure T12:** Universal Model Heterogeneity Study Scenario 7 in Table T1. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



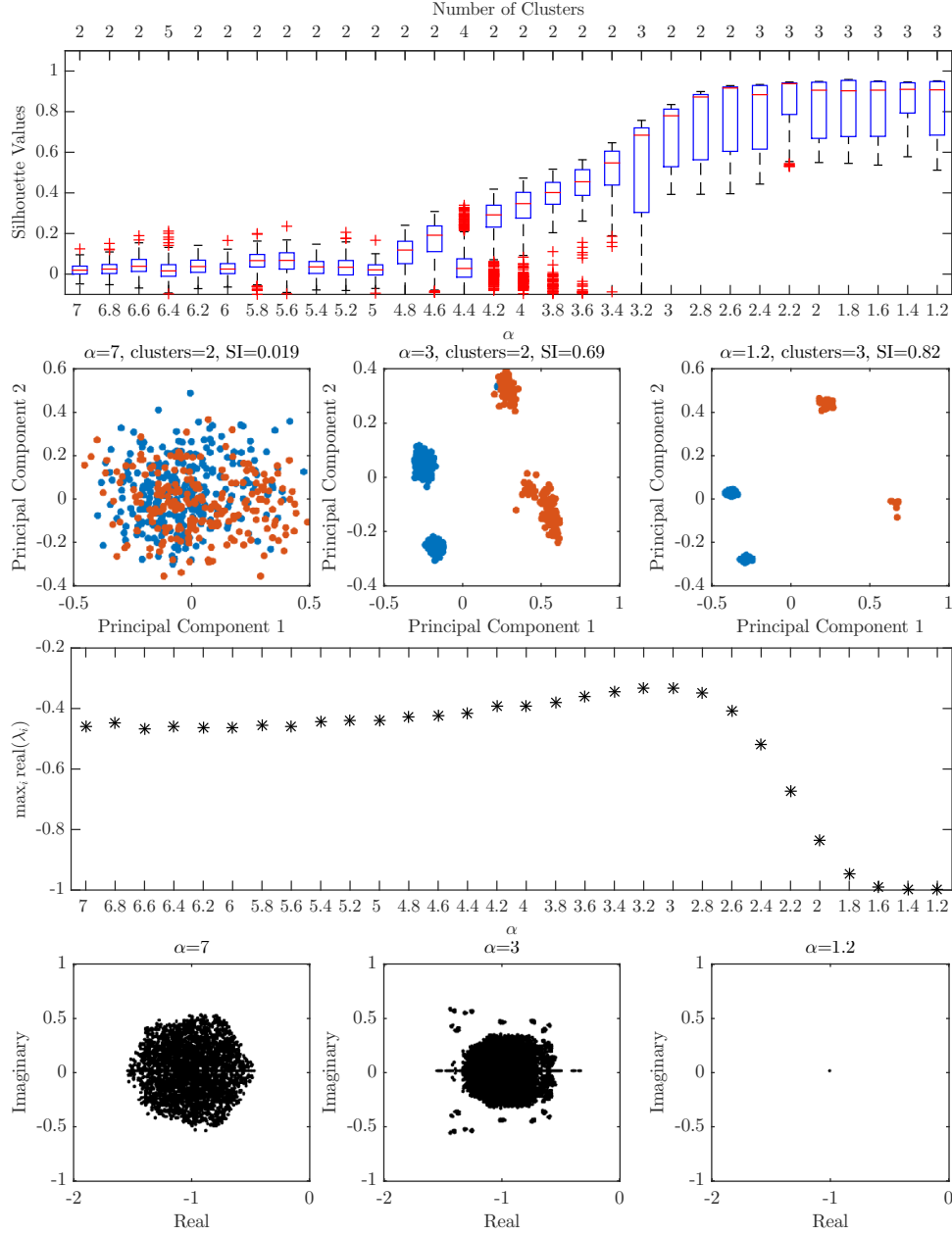
**Supplementary Text Figure T13:** Universal Model Heterogeneity Study Scenario 8 in Table T1. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



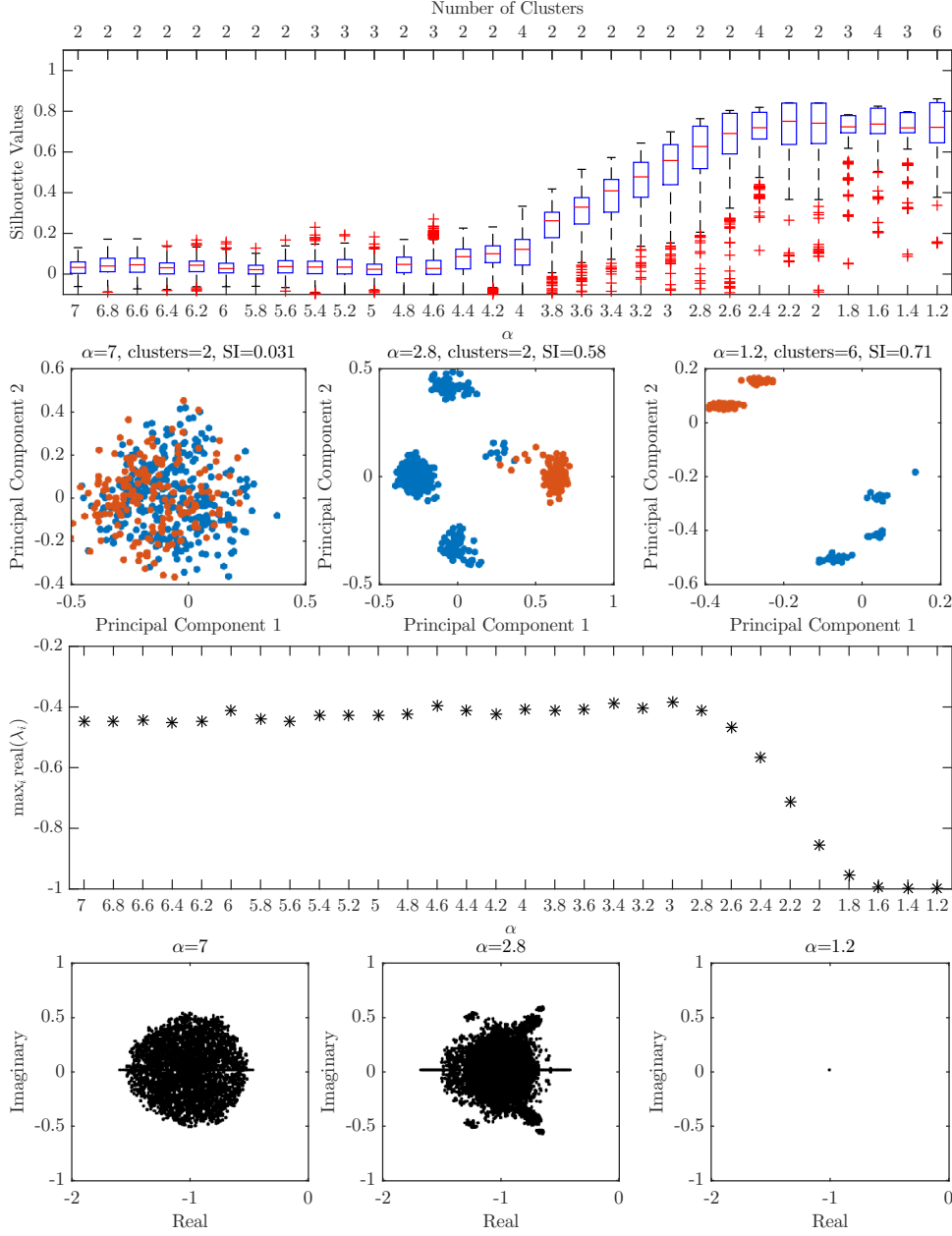
**Supplementary Text Figure T14:** Universal Model Sparsity Study Scenario 1 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row of the figure plots  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



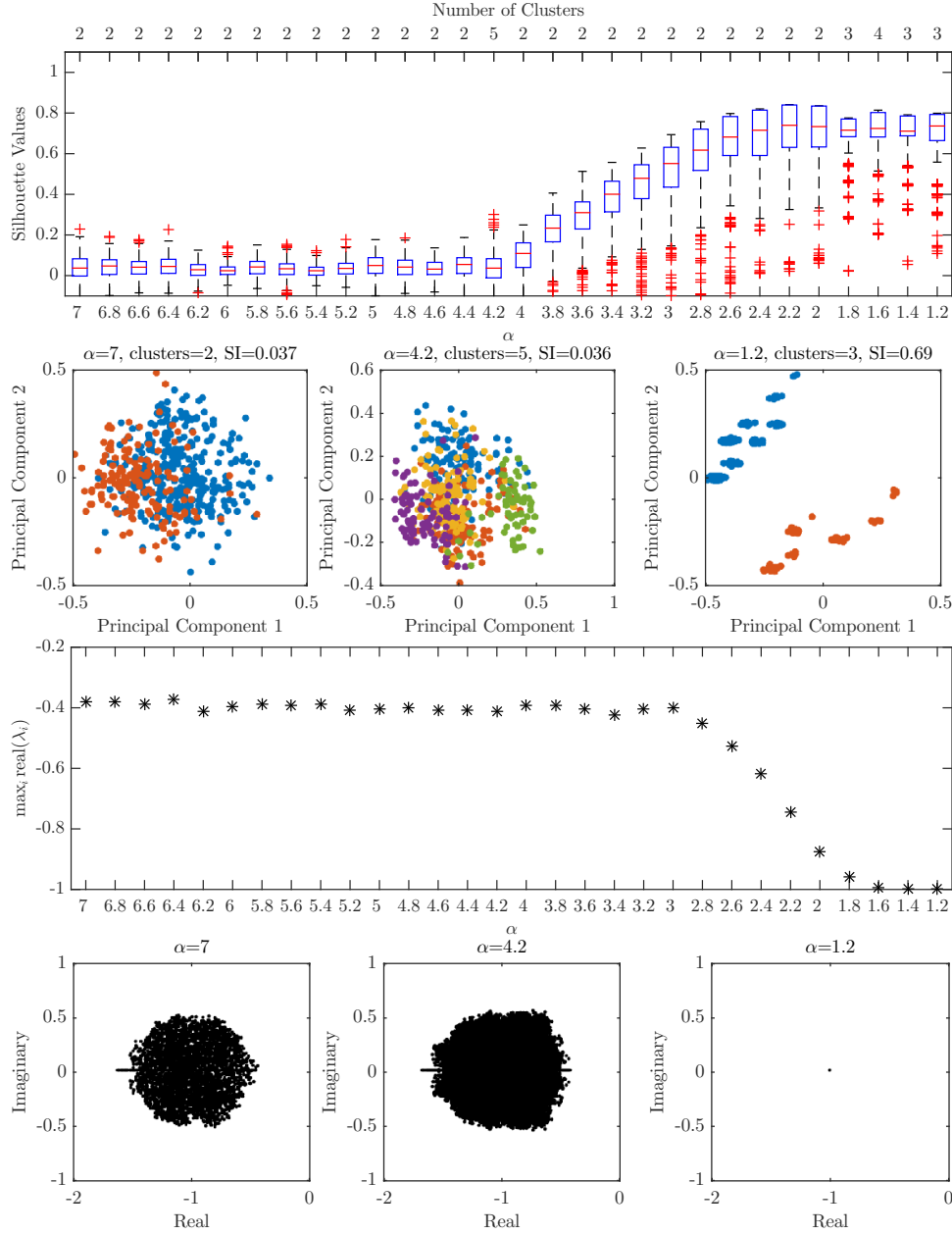
**Supplementary Text Figure T15:** Universal Model Sparsity Study Scenario 2 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



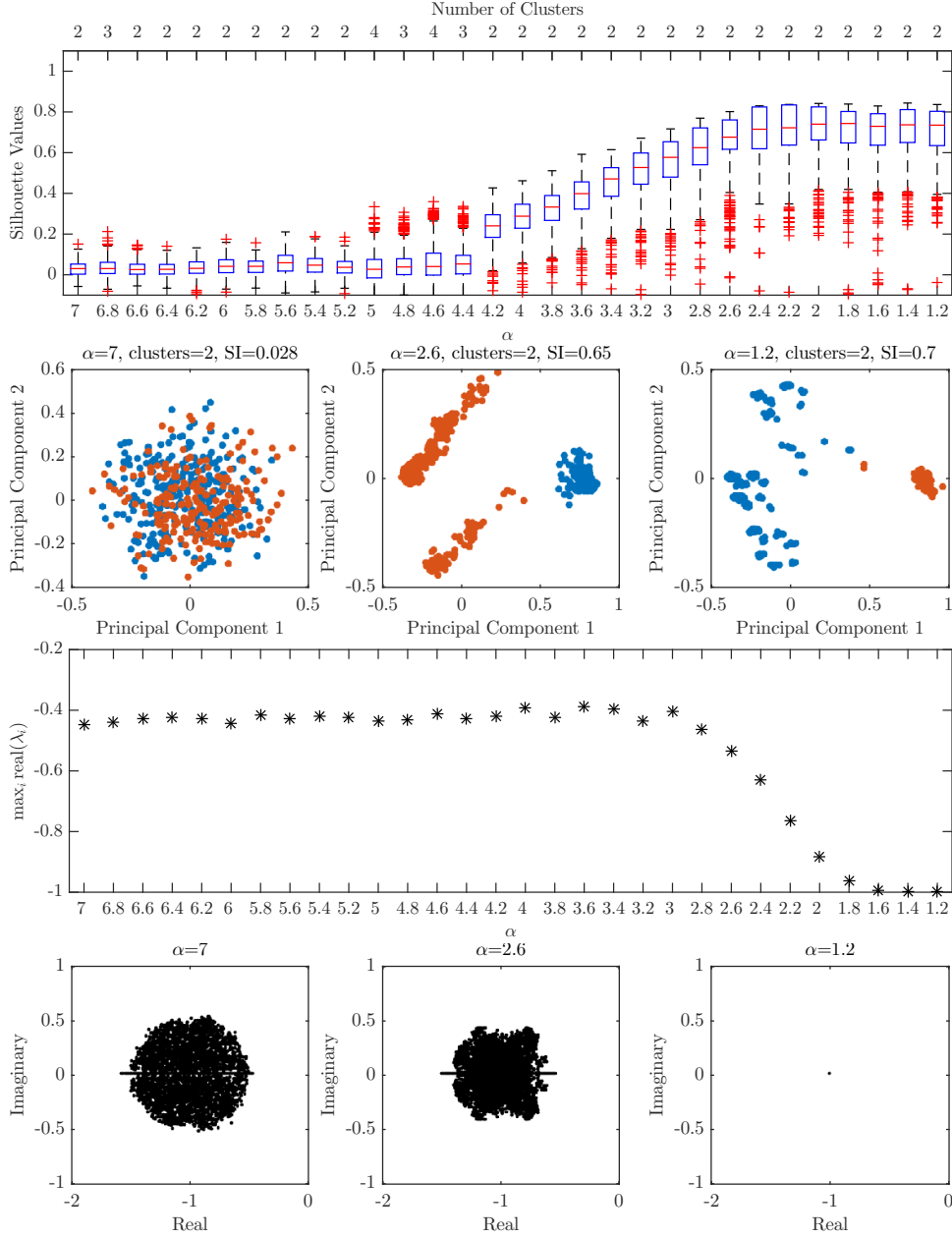
**Supplementary Text Figure T16:** Universal Model Sparsity Study Scenario 3 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \operatorname{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



**Supplementary Text Figure T17:** Universal Model Sparsity Study Scenario 4 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .

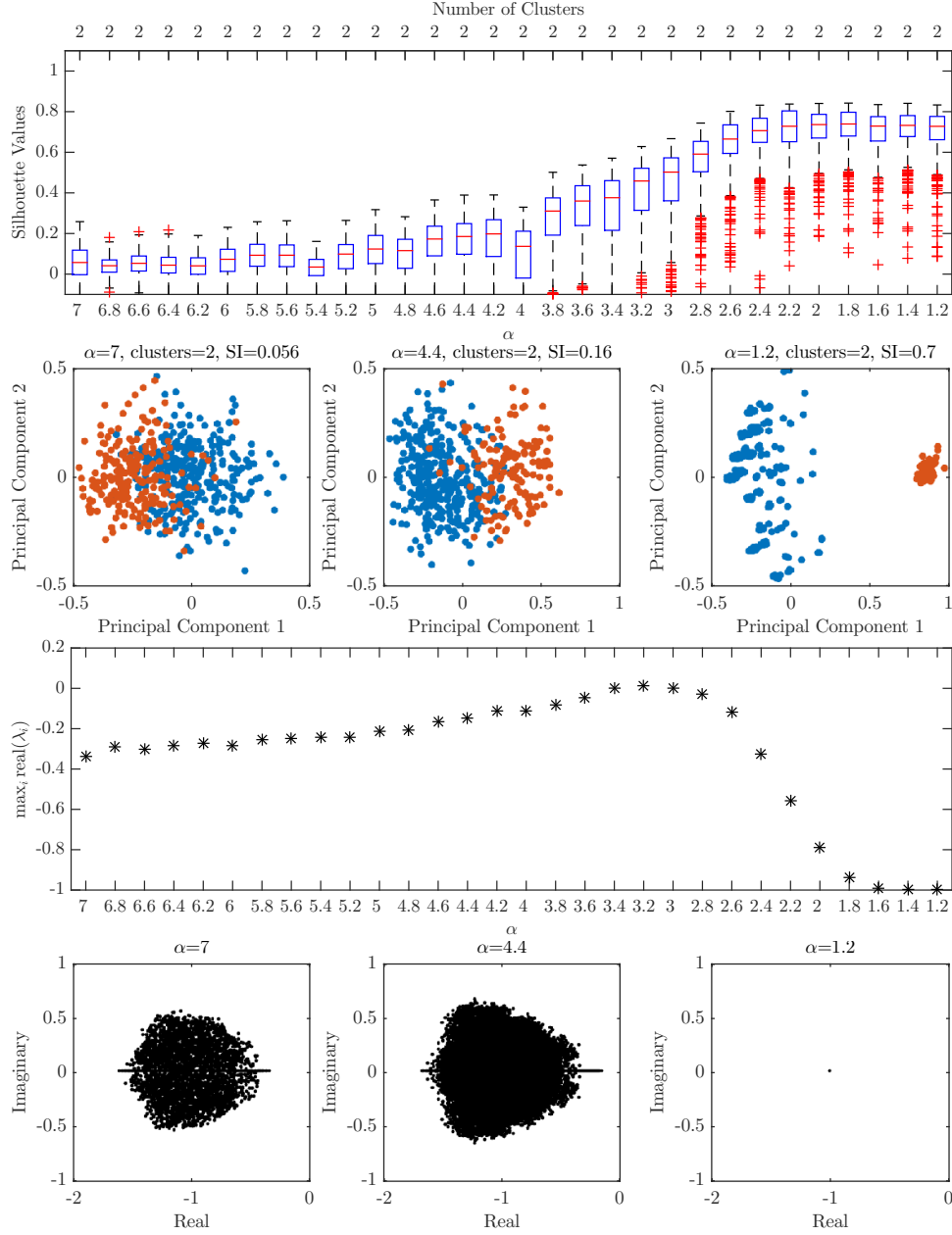


**Supplementary Text Figure T18:** Universal Model Sparsity Study Scenario 5 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row of the figure plots  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .

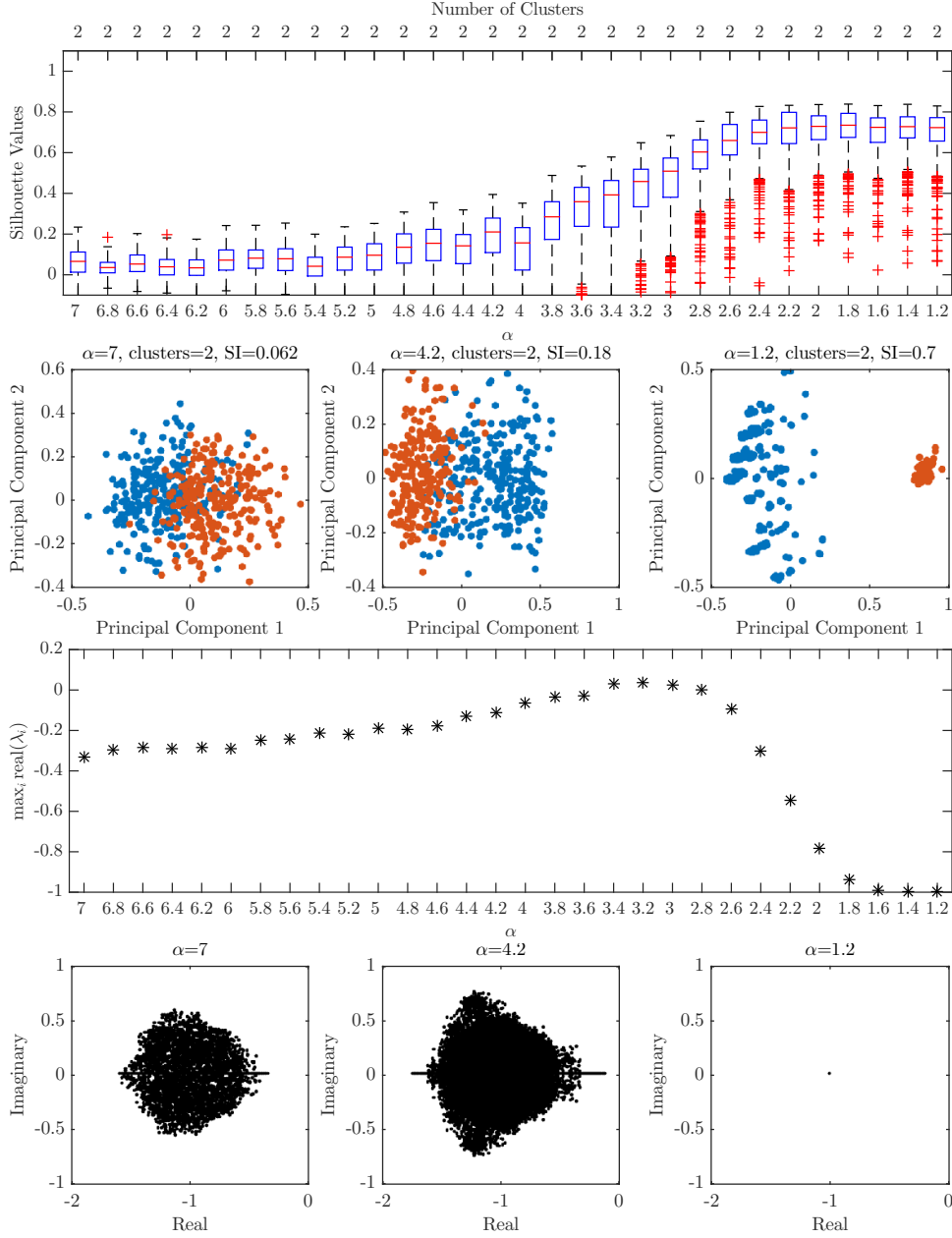


**Supplementary Text Figure T19:** Universal Model Sparsity Study Scenario 6 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .

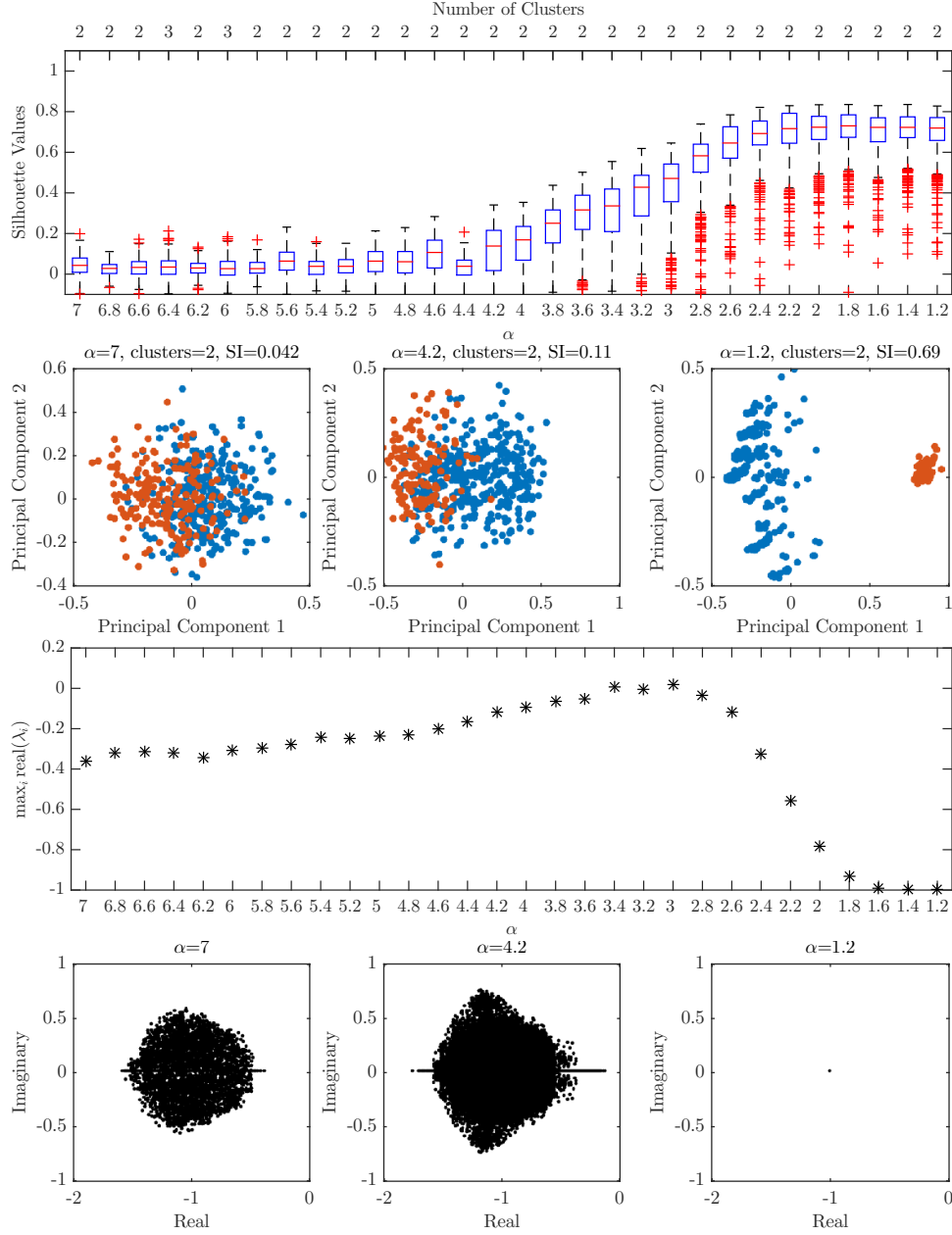




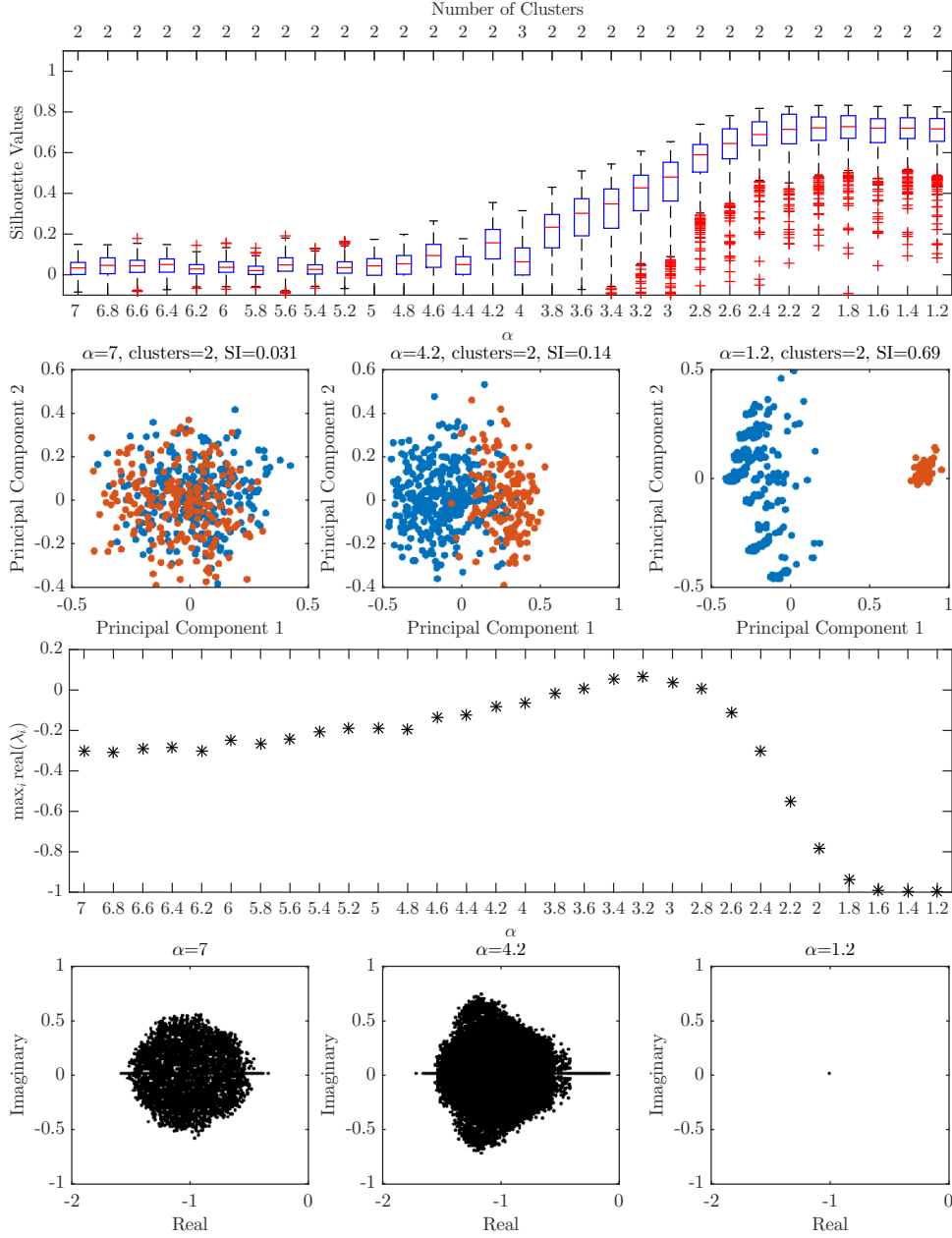
**Supplementary Text Figure T20:** Universal Model Sparsity Study Scenario 7 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_i \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



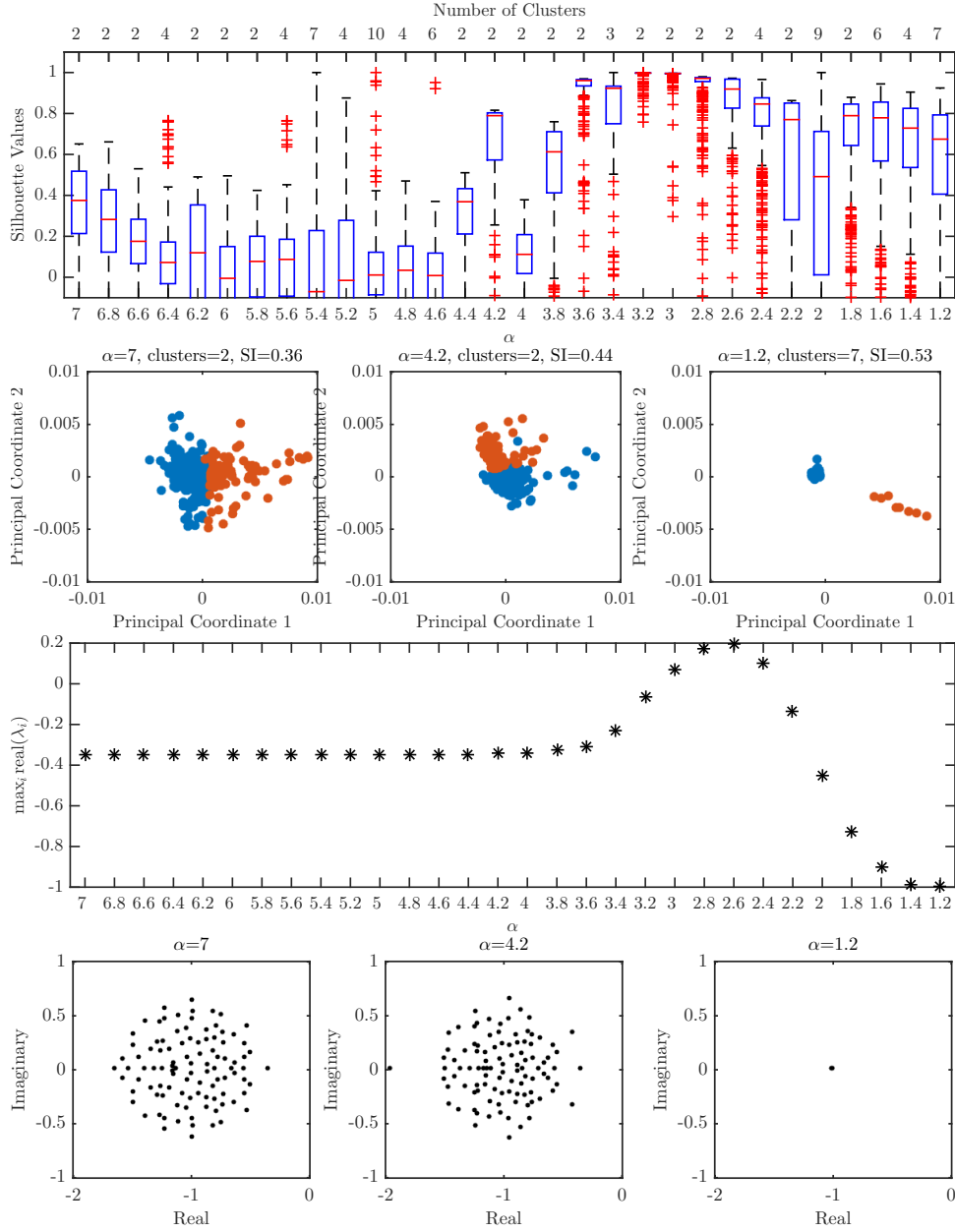
**Supplementary Text Figure T21:** Universal Model Sparsity Study Scenario 8 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



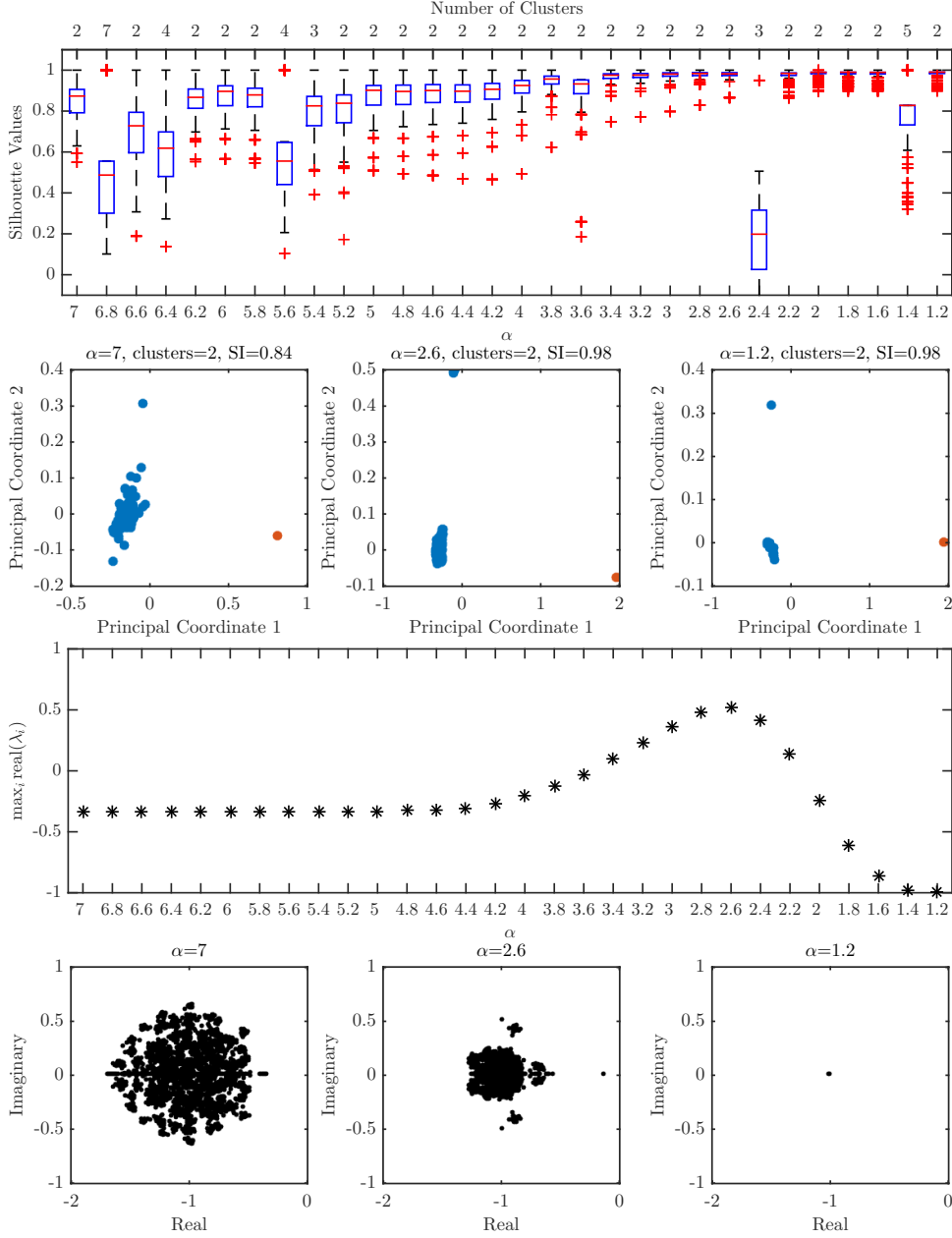
**Supplementary Text Figure T22:** Universal Model Sparsity Study Scenario 9 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



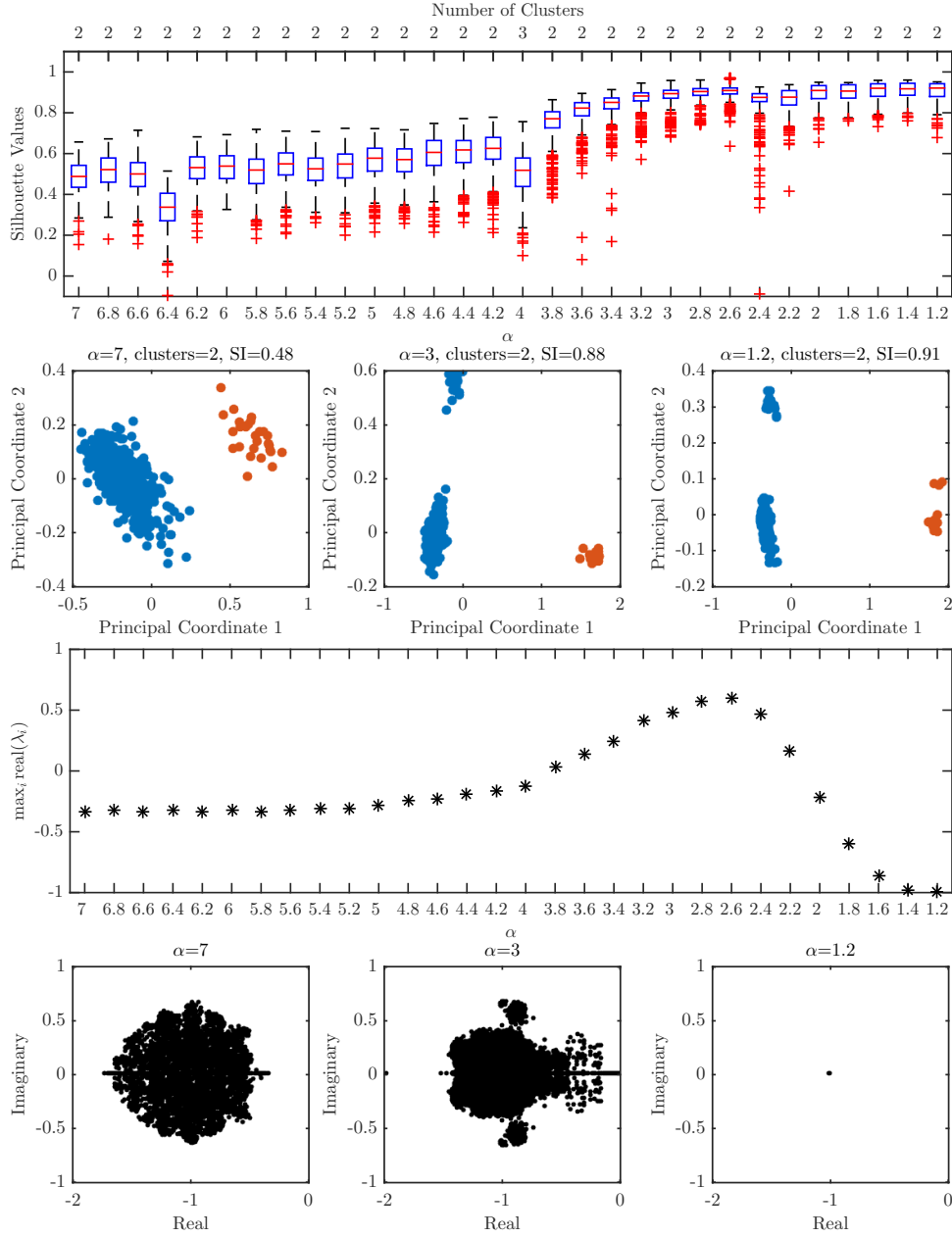
**Supplementary Text Figure T23:** Universal Model Sparsity Study Scenario 10 in Table T2. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, 80\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



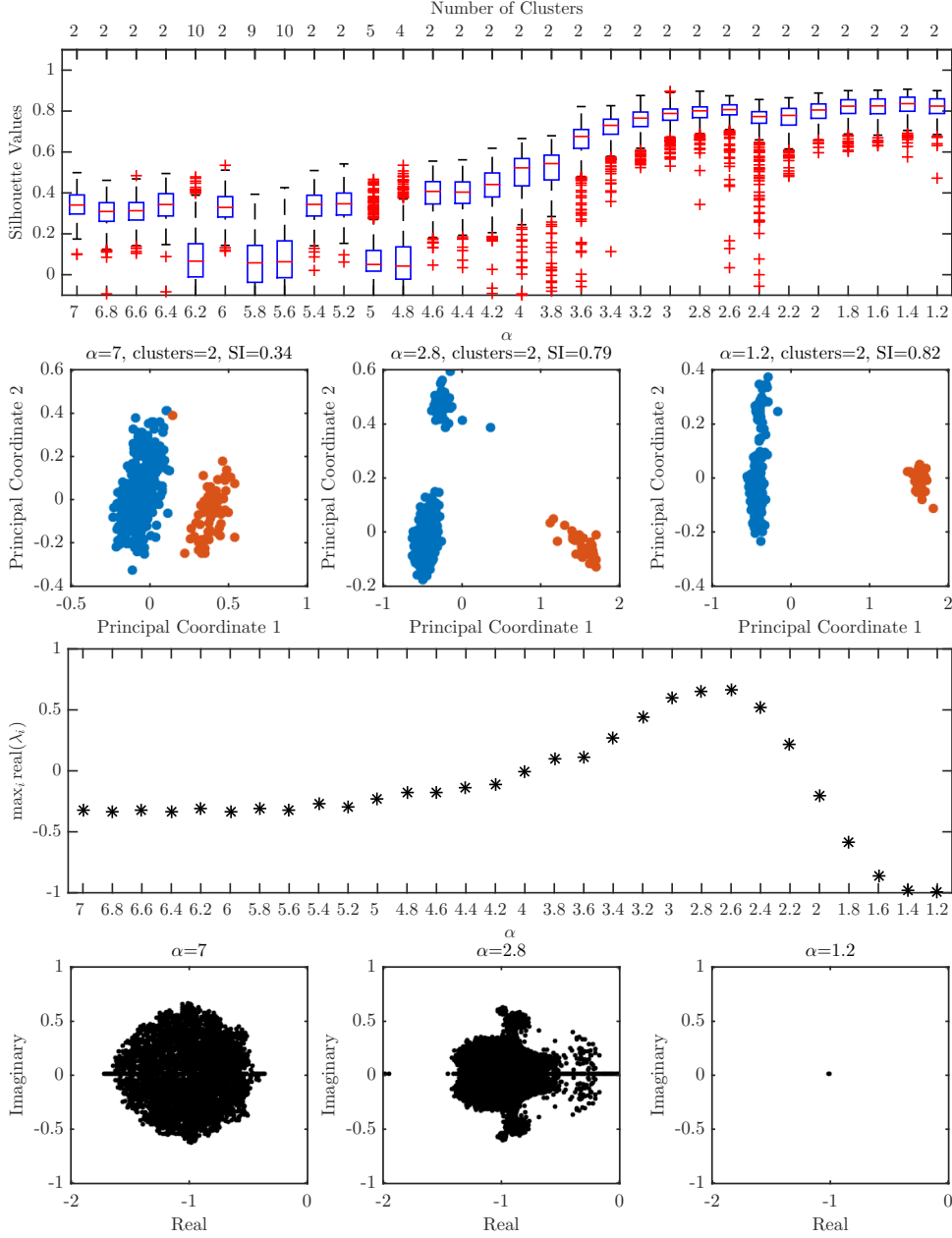
**Supplementary Text Figure T24:** Universal Model Community Size Overlap Study Scenario 1 in Table T3. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, p\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



**Supplementary Text Figure T25:** Universal Model Community Size Overlap Study Scenario 2 in Table T3. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, p\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .

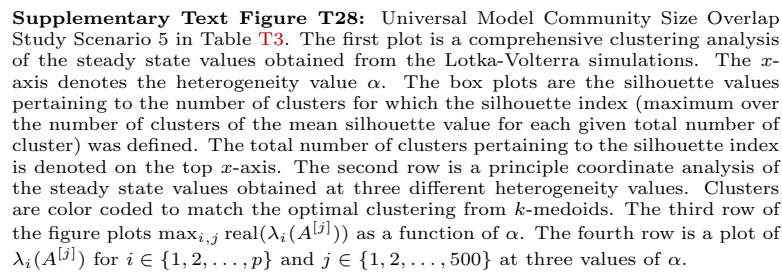


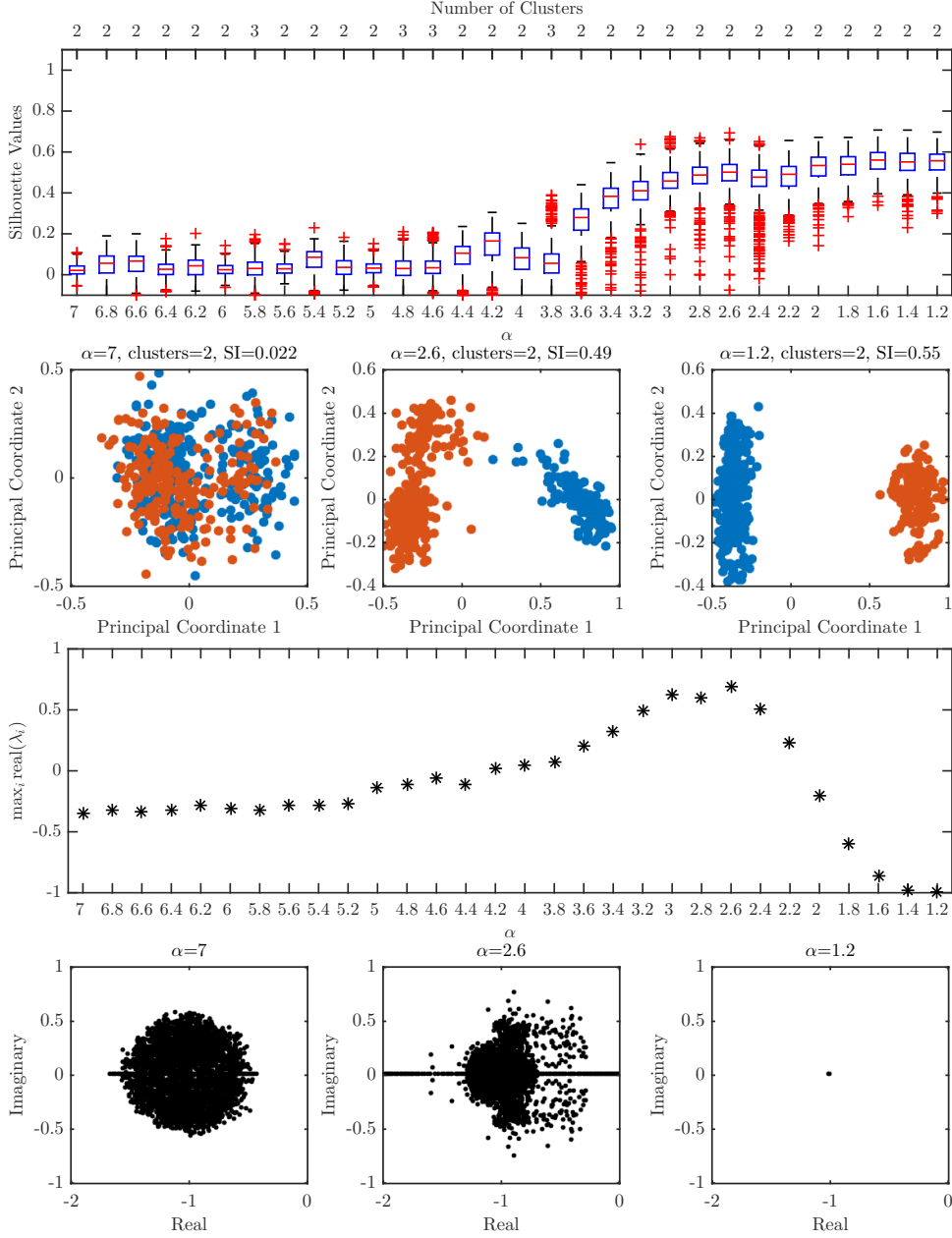
**Supplementary Text Figure T26:** Universal Model Community Size Overlap Study Scenario 3 in Table T3. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, p\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



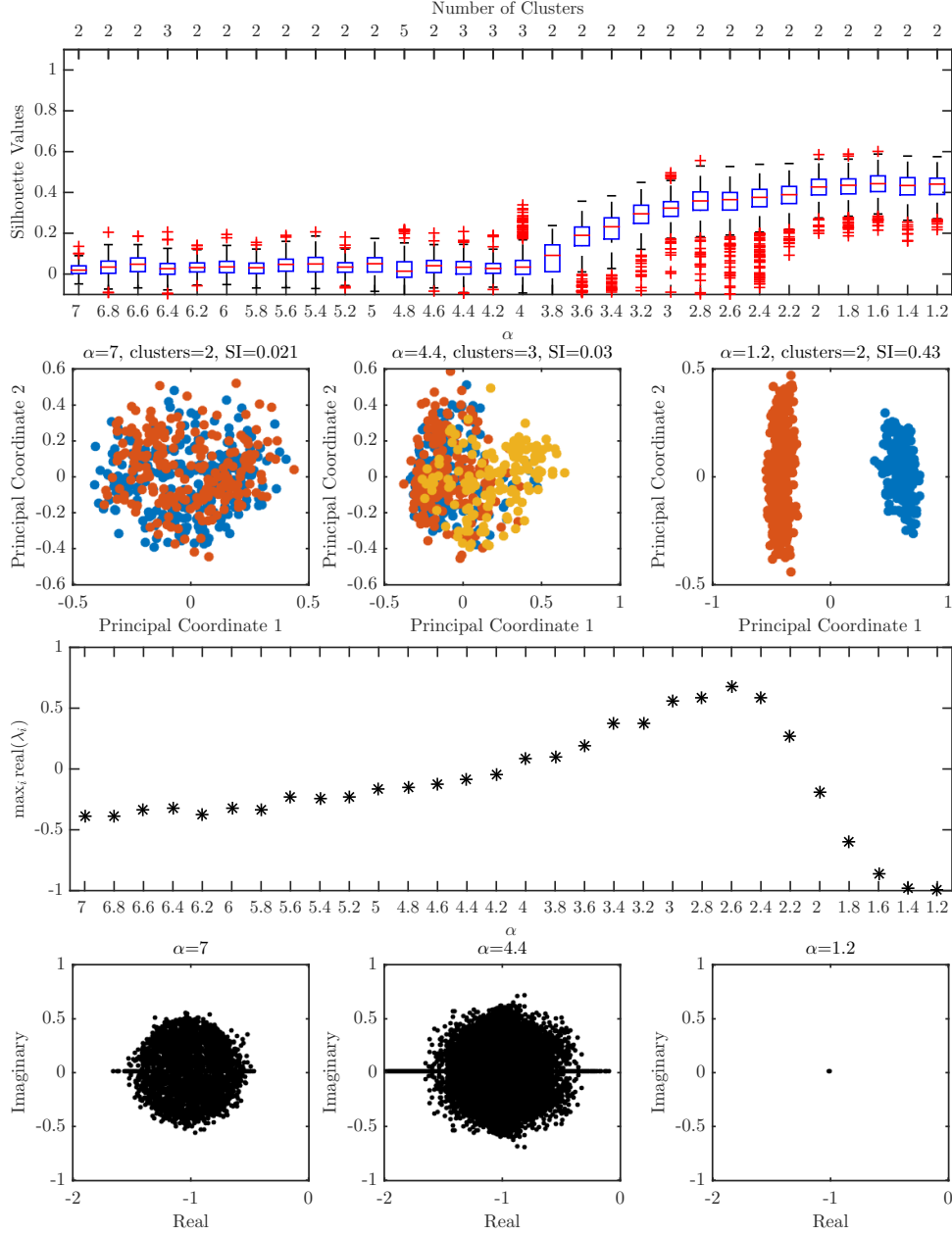
**Supplementary Text Figure T27:** Universal Model Community Size Overlap Study Scenario 4 in Table T3. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, p\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



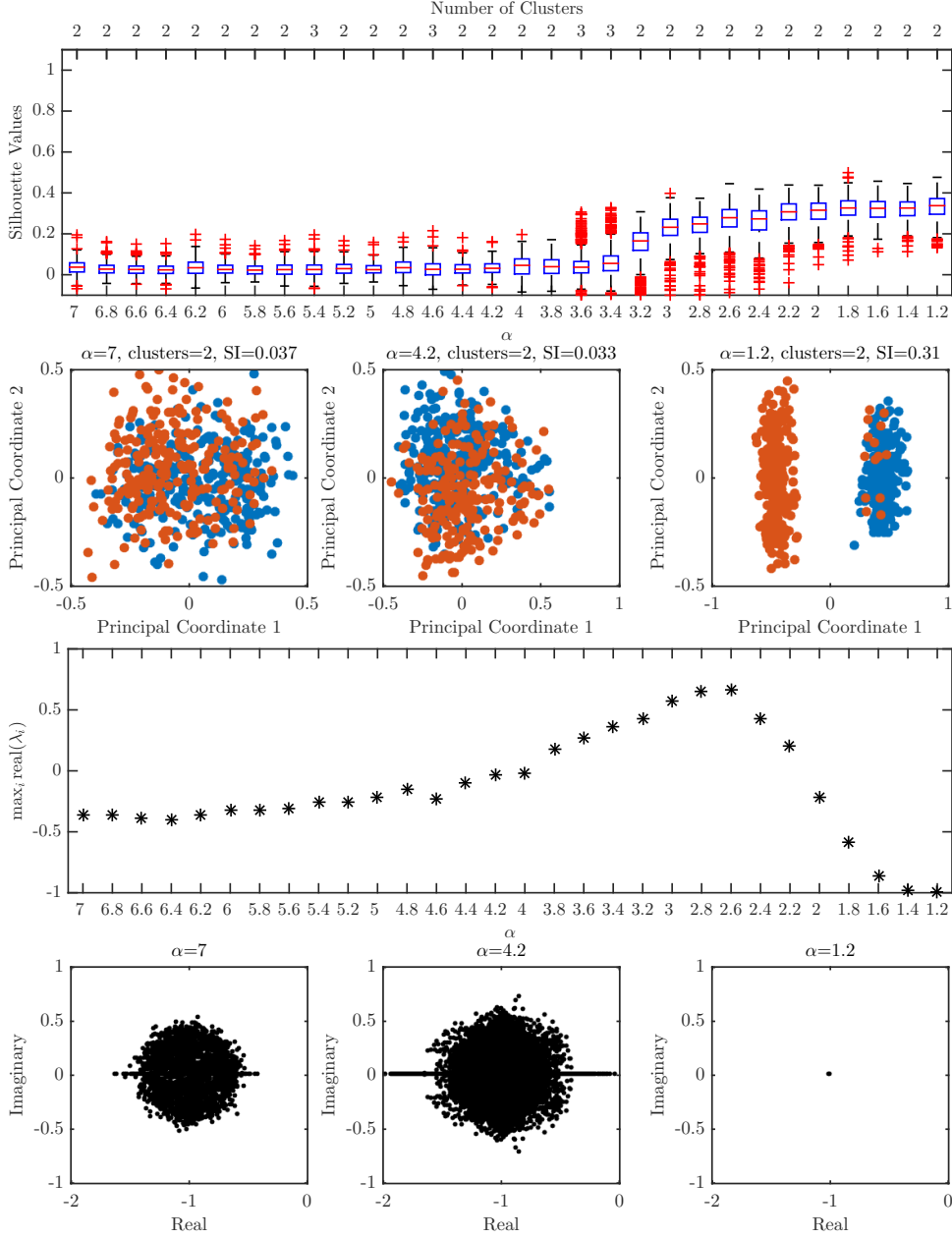




**Supplementary Text Figure T29:** Universal Model Community Size Overlap Study Scenario 6 in Table T3. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, p\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



**Supplementary Text Figure T30:** Universal Model Community Size Overlap Study Scenario 7 in Table T3. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, p\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .



**Supplementary Text Figure T31:** Universal Model Community Size Overlap Study Scenario 8 in Table T3. The first plot is a comprehensive clustering analysis of the steady state values obtained from the Lotka-Volterra simulations. The  $x$ -axis denotes the heterogeneity value  $\alpha$ . The box plots are the silhouette values pertaining to the number of clusters for which the silhouette index (maximum over the number of clusters of the mean silhouette value for each given total number of cluster) was defined. The total number of clusters pertaining to the silhouette index is denoted on the top  $x$ -axis. The second row is a principle coordinate analysis of the steady state values obtained at three different heterogeneity values. Clusters are color coded to match the optimal clustering from  $k$ -medoids. The third row of the figure plots  $\max_{i,j} \text{real}(\lambda_i(A^{[j]}))$  as a function of  $\alpha$ . The fourth row is a plot of  $\lambda_i(A^{[j]})$  for  $i \in \{1, 2, \dots, p\}$  and  $j \in \{1, 2, \dots, 500\}$  at three values of  $\alpha$ .

# Bibliography

- [1] David L Alderson and John C Doyle, *Contrasting views of complexity and their implications for network-centric infrastructures*, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on **40** (2010), no. 4, 839–852.
- [2] Marco Tulio Angulo, Jaime A Moreno, Albert-László Barabási, and Yang-Yu Liu, *Fundamental limitations of network reconstruction*, arXiv (2015), available at <http://arxiv.org/abs/1508.03559>.
- [3] Manimozhiyan Arumugam et al., *Enterotypes of the human gut microbiome*, Nature **473** (2011), no. 7346, 174–180.
- [4] ———, *Addendum: Enterotypes of the human gut microbiome*, Nature **506** (2014), no. 7489, 516–516.
- [5] ZD Bai, YQ Yin, et al., *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix*, The Annals of Probability **16** (1988), no. 4, 1729–1741.
- [6] Albert-László Barabási and Réka Albert, *Emergence of scaling in random networks*, science **286** (1999), no. 5439, 509–512.
- [7] David Berry and Stefanie Widder, *Deciphering microbial interactions and detecting keystone species with co-occurrence networks*, Frontiers in microbiology **5** (2014).
- [8] Béla Bollobás, Christian Borgs, Jennifer Chayes, and Oliver Riordan, *Directed scale-free graphs*, Proceedings of the fourteenth annual acm-siam symposium on discrete algorithms, 2003, pp. 132–139.
- [9] Béla Bollobás and Oliver M Riordan, *Mathematical results on scale-free random graphs*, Handbook of graphs and networks: from the genome to the internet (2003), 1–34.
- [10] Vanni Bucci and Joao B. Xavier, *Towards predictive models of the human gut microbiome*, Journal of Molecular Biology **426** (2014), no. 23, 3907–3916.
- [11] Charlie G. Buffie, Vanni Bucci, Richard R. Stein, Peter T. McKenney, Lilan Ling, Asia Gobourne, Daniel No, Hui Liu, Melissa Kinnebrew, Agnes Viale, Eric Littmann, Marcel R. M. van den Brink, Robert R. Jenq, Ying Taur, Chris Sander, Justin R. Cross, Nora C. Toussaint, Joao B. Xavier, and Eric G. Pamer, *Precision microbiome reconstitution restores bile acid mediated resistance to clostridium difficile*, Nature **517** (2015), no. 7533, 205–208.
- [12] T. Caliński and J Harabasz, *A dendrite method for cluster analysis*, Communications in Statistics **3** (1974), no. 1, 1–27.
- [13] J Gregory Caporaso, Christian L Lauber, Elizabeth K Costello, Donna Berg-Lyons, Antonio Gonzalez, Jesse Stombaugh, Dan Knights, Pawel Gajer, Jacques Ravel, and Noah Fierer, *Moving pictures of the human microbiome*, Genome Biology (2011).
- [14] Marcus J. Claesson, Ian B. Jeffery, Susana Conde, Susan E. Power, Eibhlís M. O’Connor, Siobhan Cusack, Hugh M. B. Harris, Mairead Coakley, Bhuvaneswari Lakshminarayanan, Orla O’Sullivan, Gerald F. Fitzgerald, Jennifer Deane, Michael O’Connor, Norma Harnedy, Kieran O’Connor, Denis O’Mahony, Douwe van Sinderen, Martina Wallace, Lorraine Brennan, Catherine Stanton, Julian R. Marchesi, Anthony P. Fitzgerald, Fergus Shanahan, Colin Hill, R. Paul Ross, and Paul W. O’Toole, *Gut microbiota composition correlates with diet and health in the elderly*, Nature **488** (201208), no. 7410, 178–184.
- [15] A. Clauset, C. Shalizi, and M. Newman, *Power-law distributions in empirical data*, SIAM Review **51** (2009), no. 4, 661–703.
- [16] Earl A Coddington and Norman Levinson, *Theory of ordinary differential equations*, Tata McGraw-Hill Education, 1955.
- [17] Joseph H Connell and Wayne P Sousa, *On the evidence needed to judge ecological stability or persistence*, American Naturalist (1983), 789–824.
- [18] Elizabeth K. Costello, Keaton Stagaman, Les Dethlefsen, Brendan J. M. Bohannon, and David A. Relman, *The application of ecological theory toward an understanding of the human microbiome*, Science **336** (2012), no. 6086, 1255–1262.
- [19] G. W. Cross, *Three types of matrix stability*, Linear Algebra and its Applications **20** (19786), no. 3, 253–263.
- [20] Lawrence A David, Arne C Materna, Jonathan Friedman, Maria I Campos-Baptista, Matthew C Blackburn, Allison Perrotta, Susan E Erdman, and Eric J Alm, *Host lifestyle affects human microbiota on daily timescales*, Genome Biology (2014).
- [21] Hua Deng, Miroslav Krstić, and Ruth J Williams, *Stabilization of stochastic nonlinear systems driven by noise of unknown covariance*, Automatic Control, IEEE Transactions on **46** (2001), no. 8, 1237–1253.

- [22] Tao Ding and Patrick D. Schloss, *Dynamics and associations of microbial community types across the human body*, Nature **509** (201405), no. 7500, 357–360.
- [23] Joseph L Doob, *Stochastic processes*, Wiley, 1990.
- [24] ———, *Classical potential theory and its probabilistic counterpart*, Springer Science & Business Media, 2012.
- [25] Andrew W. DuPont and Herbert L. DuPont, *The intestinal microbiota and chronic disorders of the gut*, Nat Rev Gastroenterol Hepatol **8** (201109), no. 9, 523–531.
- [26] Anna Eklöf and BO Ebenman, *Species loss and secondary extinctions in simple and complex model communities*, Journal of Animal Ecology **75** (2006), no. 1, 239–246.
- [27] Mark C. Emmerson and Dave Raffaelli, *Predator–prey body size, interaction strength and the stability of a real food web*, Journal of Animal Ecology **73** (2004), no. 3, 399–409.
- [28] Erdős and Rényi, *On random graphs i*, Publicationes Mathematicae Debrecen **6** (1959), 290–297.
- [29] ———, *On the evolution of random graphs*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5** (1960), 17–61.
- [30] Walter R. Evans, *Control system synthesis by root locus method*, American Institute of Electrical Engineers, Transactions of the **69** (1950), no. 1, 66–69.
- [31] Jeremiah J. Faith, Janaki L. Gurge, Mark Charbonneau, Sathish Subramanian, Henning Seedorf, Andrew L. Goodman, Jose C. Clemente, Rob Knight, Andrew C. Heath, Rudolph L. Leibel, Michael Rosenbaum, and Jeffrey I. Gordon, *The long-term stability of the human gut microbiota*, Science **341** (2013), no. 6141.
- [32] Karoline Faust and Jeroen Raes, *Microbial interactions: from networks to models*, Nature Reviews Microbiology **10** (2012), no. 8, 538–550.
- [33] Charles K. Fisher and Pankaj Mehta, *Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression*, PLoS ONE **9** (2014), no. 7, e102451.
- [34] ———, *The transition between the niche and neutral regimes in ecology*, Proceedings of the National Academy of Sciences **111** (2014), no. 36, 13111–13116.
- [35] Georg K. Gerber, *The dynamic microbiome*, FEBS Letters **588** (201411), no. 22, 4131–4139.
- [36] E. N. Gilbert, *Random graphs*, The Annals of Mathematical Statistics **30** (1959), no. 4, 1141–1144.
- [37] Narendra S Goel, Samaresh C Maitra, and Elliott W Montroll, *On the volterra and other nonlinear models of interacting populations*, Reviews of modern physics **43** (1971), no. 2, 231.
- [38] B. S. Goh, *Global stability in many-species systems*, American Naturalist (1977), 135–143.
- [39] Julia K. Goodrich, Sara C. Di Rienzi, Angela C. Poole, Omry Koren, William A. Walters, J. Gregory Caporaso, Rob Knight, and Ruth E. Ley, *Conducting a microbiome study*, Cell **158** (20147), no. 2, 250–262.
- [40] W. Hahn, *Stability of motion*, Springer-Verlag, New York NY, 1967.
- [41] Falk Hildebrand, TL Nguyen, Brigitta Brinkman, Roberto Garcia Yunta, Benedicte Cauwe, Peter Vandenabeele, Adrian Liston, and Jeroen Raes, *Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice*, Genome Biology (2013).
- [42] Kiyosi Itô, *Stochastic integral*, Proc. Imperial Acad. Tokyo **20** (1944), 519–524.
- [43] Kiyosi Itô, *On a stochastic integral equation*, Proceedings of the Japan Academy **22** (1946), no. 1-4, 32–35.
- [44] A. K. Jain, M. N. Murty, and P. J. Flynn, *Data clustering: A review*, ACM Comput. Surv. **31** (1999), no. 3, 264–323.
- [45] Ian B. Jeffery, Marcus J. Claesson, Paul W. O’Toole, and Fergus Shanahan, *Categorization of the gut microbiota: enterotypes or gradients?*, Nat Rev Micro **10** (201209), no. 9, 591–592.
- [46] H. Jeffreys, *The theory of probability*, 2nd ed., Oxford, 1948.
- [47] Johan Ludwig William Valdemar Jensen, *Sur les fonctions convexes et les inégalités entre les valeurs moyennes*, Acta Mathematica **30** (1906), no. 1, 175–193.
- [48] R. E. Kalman and J. E. Bertram, *Control systems analysis and design via the ‘second method’ of liapunov, i. continuous-time systems*, Journal of Basic Engineering **82** (1960), 371–393.
- [49] Leonard Kaufman and Peter Rousseeuw, *Clustering by means of medoids*, North-Holland, 1987.
- [50] Rafail Khasminskii, *Stochastic stability of differential equations*, Vol. 66, Springer Science & Business Media, 2011.
- [51] Dan Knights, Tonya L. Ward, Christopher E. McKinlay, Hannah Miller, Antonio Gonzalez, Daniel McDonald, and Rob Knight, *Rethinking “enterotypes”*, Cell Host & Microbe **16** (2014), no. 4, 433–437.
- [52] Omry Koren, Dan Knights, Antonio Gonzalez, Levi Waldron, Nicola Segata, Rob Knight, Curtis Huttenhower, and Ruth E. Ley, *A guide to enterotypes across the human body: Meta-analysis of microbial community structures in human microbiome datasets*, PLoS Comput Biol **9** (2013), no. 1.
- [53] Jonathan W. Kotula, S. Jordan Kerns, Lev A. Shaket, Layla Siraj, James J. Collins, Jeffrey C. Way, and Pamela A. Silver, *Programmable bacteria detect and record an environmental signal in the mammalian gut*, Proceedings of the National Academy of Sciences **111** (2014), no. 13, 4838–4843.

- [54] Solomon Kullback and Richard A Leibler, *On information and sufficiency*, The Annals of Mathematical Statistics (1951), 79–86.
- [55] Effendi Leonard, David Nielsen, Kevin Solomon, and Kristala Jones Prather, *Engineering microbes with synthetic biology frameworks*, Trends in biotechnology **26** (2008), no. 12, 674–681.
- [56] Roie Levy and Elhanan Borenstein, *Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules*, Proceedings of the National Academy of Sciences **110** (2013), no. 31, 12804–12809.
- [57] Richard C Lewontin, *The meaning of stability*, Brookhaven symposia in biology, 1969, pp. 13.
- [58] Lun Li, David Alderson, John C Doyle, and Walter Willinger, *Towards a theory of scale-free graphs: Definition, properties, and implications*, Internet Mathematics **2** (2004), no. 4, 431–523.
- [59] Jianhua Lin, *Divergence measures based on the shannon entropy*, Information Theory, IEEE Transactions on **37** (1991Jan), no. 1, 145–151.
- [60] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási, *Controllability of complex networks*, Nature **473** (2011), no. 7346, 167–173.
- [61] Catherine A. Lozupone, Jesse I. Stombaugh, Jeffrey I. Gordon, Janet K. Jansson, and Rob Knight, *Diversity, stability and resilience of the human gut microbiota*, Nature **489** (201209), no. 7415, 220–230.
- [62] Simeone Marino, Nielson T. Baxter, Gary B. Huffnagle, Joseph F. Petrosino, and Patrick D. Schloss, *Mathematical modeling of primary succession of murine intestinal microbiota*, Proceedings of the National Academy of Sciences **111** (2014), no. 1, 439–444.
- [63] J. S. Massera, *Contributions to stability theory*, Annals of Mathematics **64** (1956 (Erratum, Ann. of Math, 1958)), no. 1.
- [64] Kevin McCann, Alan Hastings, and Gary R. Huxel, *Weak trophic interactions and the balance of nature*, Nature **395** (199810), no. 6704, 794–798.
- [65] Michael T Mee, James J Collins, George M Church, and Harris H Wang, *Syntrophic exchange in synthetic microbial communities*, Proceedings of the National Academy of Sciences **111** (2014), no. 20, E2149–E2156.
- [66] Paul André Meyer, *Probability and potentials*, Vol. 7, Blaisdell Publishing Company Waltham (Mass.), 1966.
- [67] Glenn W Milligan and Martha C Cooper, *Methodology review: Clustering methods*, Applied psychological measurement **11** (1987), no. 4, 329–354.
- [68] Rafael Muñoz-Tamayo, Béatrice Laroche, Éric Walter, Joël Doré, and Marion Leclerc, *Mathematical modelling of carbohydrate degradation by human colonic microbiota*, Journal of Theoretical Biology **266** (20109), no. 1, 189–201.
- [69] K. S. Narendra and A. M. Annaswamy, *Robust adaptive control in the presence of bounded disturbances*, IEEE Trans. Automat. Contr. **31** (1986), 306–315.
- [70] ———, *Stable adaptive systems*, Dover, 2005.
- [71] Karen E. Nelson and S. Dusko Ehrlich, *Metahit: The european union project on metagenomics of the human intestinal tract*, Springer New York, 2011.
- [72] Takashi Nishikawa, Adilson E Motter, Ying-Cheng Lai, and Frank C Hoppensteadt, *Heterogeneity in oscillator networks: Are smaller worlds easier to synchronize?*, Physical review letters **91** (2003), no. 1, 014101.
- [73] R. T. Paine, *Food-web analysis through field measurement of per capita interaction strength*, Nature **355** (199201), no. 6355, 73–75.
- [74] Robert T Paine, *A conversation on refining the concept of keystone species*, JSTOR, 1995.
- [75] Romualdo Pastor-Satorras and Alessandro Vespignani, *Epidemic spreading in scale-free networks*, Physical review letters **86** (2001), no. 14, 3200.
- [76] John W Pepper and Simon Rosenfeld, *The emerging medical ecology of the human gut microbiome*, Trends in ecology & evolution **27** (2012), no. 7, 381–384.
- [77] William H Press, Brian P Flannery, Saul A Teukolsky, and William T Vetterling, *Numerical recipes: The art of scientific computing*, 3rd ed., Cambridge University Press London, 2007.
- [78] Philip E Protter, *Stochastic integration and differential equations*, Vol. 21, Springer, 2013.
- [79] Jacques Ravel, Pawel Gajer, Zaid Abdo, G. Maria Schneider, Sara S. K. Koenig, Stacey L. McCulle, Shara Karlebach, Reshma Gorle, Jennifer Russell, Carol O. Tacket, Rebecca M. Brotman, Catherine C. Davis, Kevin Ault, Ligia Peralta, and Larry J. Forney, *Vaginal microbiome of reproductive-age women*, Proceedings of the National Academy of Sciences **108** (2011), 4680–4687.
- [80] David A Relman, *The human microbiome: ecosystem resilience and health*, Nutrition reviews **70** (2012), no. suppl 1, S2–S9.
- [81] June L. Round and Sarkis K. Mazmanian, *The gut microbiota shapes intestinal immune responses during health and disease*, Nat Rev Immunol **9** (200905), no. 5, 313–323.
- [82] George AF Seber, *Multivariate observations*, John Wiley & Sons, 1984.
- [83] Saeed Shoaie, Fredrik Karlsson, Adil Mardinoglu, Intawat Nookaew, Sergio Bordel, and Jens Nielsen, *Understanding the interactions between bacteria in the human gut through metabolic modeling*, Sci. Rep. **3** (201308).
- [84] Robert Shorten and Kumpati S Narendra, *On a theorem of redheffer concerning diagonal stability*, Linear Algebra and its Applications **431** (2009), no. 12, 2317–2329.

- [85] Justin L. Sonnenburg, *Engineering the human microbiome shows promise for treating disease*, Scientific American **312** (2015), no. 3.
- [86] Richard R. Stein, Vanni Bucci, Nora C. Toussaint, Charlie G. Buffie, Gunnar Rättsch, Eric G. Pamer, Chris Sander, and João B. Xavier, *Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota*, PLoS Comput Biol **9** (2013), no. 12.
- [87] B.L. Stevens and F.L. Lewis, *Aircraft control and simulation*, Wiley, 2003.
- [88] Daniel W Stroock, *Probability theory: an analytic view*, Cambridge university press, 2010.
- [89] Yu Tanouchi, Robert P Smith, and Lingchong You, *Engineering microbial systems to explore ecological and evolutionary dynamics*, Current opinion in biotechnology **23** (2012), no. 5, 791–797.
- [90] T. Tao and V. Vu, *Random matrices: The Universality phenomenon for Wigner ensembles*, ArXiv (January 2012), available at <http://arxiv.org/abs/1202.0068>.
- [91] Terence Tao, *An introduction to measure theory*, Vol. 126, American Mathematical Soc., 2011.
- [92] ———, *Topics in random matrix theory*, Vol. 132, American Mathematical Soc., 2012.
- [93] Terence Tao, Van Vu, and Manjunath Krishnapur, *Random matrices: Universality of esds and the circular law*, The Annals of Probability **38** (201009), no. 5, 2023–2065.
- [94] The Human Microbiome Project Consortium, *A framework for human microbiome research*, Nature **486** (2012), no. 7402, 215–221.
- [95] ———, *Structure, function and diversity of the healthy human microbiome*, Nature **486** (2012), no. 7402, 207–214.
- [96] Andrey Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Soviet math. dokl., 1963, pp. 1035–1038.
- [97] Alison Waldram, Elaine Holmes, Yulan Wang, Mattias Rantalainen, Ian D Wilson, Kieran M Tuohy, Anne L McCartney, Glenn R Gibson, and Jeremy K Nicholson, *Top-down systems biology modeling of host metabotype- microbiome associations in obese rodents*, Journal of proteome research **8** (2009), no. 5, 2361–2375.
- [98] Alexa Weingarden, Antonio González, Yoshiki Vázquez-Baeza, Sophie Weiss, Gregory Humphry, Donna Berg-Lyons, Dan Knights, Tatsuya Unno, Aleh Bobr, Johnthomas Kang, Alexander Khodruts, Rob Knight, and Michael J Sadowsky, *Dynamic changes in short-and long-term bacterial composition following fecal microbiota transplantation for recurrent clostridium difficile infection*, Microbiome **3** (2015), no. 1, 10.
- [99] Eugene P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics **62** (1955/11/01), no. 3, 548–564.
- [100] ———, *Characteristic vectors of bordered matrices with infinite dimensions ii*, Annals of Mathematics **65** (1957/03/01), no. 2, 203–207.
- [101] ———, *On the distribution of the roots of certain symmetric matrices*, Annals of Mathematics **67** (1958/03/01), no. 2, 325–327.
- [102] Gary D. Wu, Jun Chen, Christian Hoffmann, Kyle Bittinger, Ying-Yu Chen, Sue A. Keilbaugh, Meenakshi Bewtra, Dan Knights, William A. Walters, Rob Knight, Rohini Sinha, Erin Gilroy, Kernika Gupta, Robert Baldassano, Lisa Nessel, Hongzhe Li, Frederic D. Bushman, and James D. Lewis, *Linking long-term dietary patterns with gut microbial enterotypes*, Science **334** (2011), no. 6052, 105–108.
- [103] Stephanie J Yaung, George M Church, and Harris H Wang, *Recent progress in engineering human-associated microbiomes*, Engineering and analyzing multicellular systems, 2014, pp. 3–25.
- [104] Stephanie J Yaung, Luxue Deng, Ning Li, Jonathan L Braff, George M Church, Lynn Bry, Harris H Wang, and Georg K Gerber, *Improving microbial fitness in the mammalian gut by in vivo temporal functional metagenomics*, Molecular systems biology **11** (2015), no. 3.
- [105] I Youngster, GH Russell, C Pindar, T Ziv-Baran, J Sauk, and EL Hohmann, *Oral, capsulized, frozen fecal microbiota transplantation for relapsing clostridium difficile infection*, JAMA **312** (2014), no. 17, 1772–1778.
- [106] Ilan Youngster, Jenny Sauk, Christina Pindar, Robin G. Wilson, Jess L. Kaplan, Mark B. Smith, Eric J. Alm, Dirk Gevers, George H. Russell, and Elizabeth L. Hohmann, *Fecal microbiota transplant for relapsing clostridium difficile infection using a frozen inoculum from unrelated donors: A randomized, open-label, controlled pilot study*, Clinical Infectious Diseases **58** (2014), no. 11, 1515–1522.
- [107] Yanjiao Zhou, Kathie A Mihindukulasuriya, Hongyu Gao, Patricio S La Rosa, Kristine M Wylie, John C Martin, Karthik Kota, William D Shannon, Makedonka Mitreva, Erica Sodergren, and George M Weinstock, *Exploration of bacterial community classes in major human habitats*, Genome Biology (2014).